

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 08:09:23 ; Search time 3973.1 Seconds

(without alignments)
19326.012 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaaagtcacgctgtcctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: OLIGO_NDC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST.*
1: em_estdb.*
2: em_esthum.*
3: em_estlin.*
4: em_estmu.*
5: em_estov.*
6: em_estro.*
7: em_estro.*
8: em_hlc.*
9: gb_estl.*
10: gb_est2.*
11: gb_hlc.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683	12.0	767	10 B1917860	B1917860 603183990
2	625	11.0	725	10 BE792557	BE792557 601585341
3	589	10.4	1150	10 BM466811	BM466811 AGENCOURT
4	582	10.2	870	10 BG743714	BG743714 602632895
5	563	9.9	713	10 BG170486	BG170486 602322859
6	536	9.4	536	9 A1079538	A1079538 oz04e06.x
7	536	9.4	536	9 A1497808	A1497808 tm89f02.x
8	536	9.4	561	9 A1038061	A1038061 ox21b05.x
9	530	9.3	608	9 AM452651	AM452651 UI-H-B13-
10	523	9.2	592	9 A1599694	A1599694 DKFDP313F
11	522	9.2	574	10 BG470230	BG470230 602533745
12	517	9.1	645	10 BE693704	BE693704 RCO-BT038
13	516	9.1	509	9 AM780417	AM780417 ho21b08.x
14	492	8.6	899	10 B1093826	B1093826 602860606
15	487	8.6	575	10 B1870563	B1870563 603394044
16	486	8.5	486	9 A1952086	A1952086 wx46g11.x
17	484	8.5	486	9 A1333325	A1333325 q905b01.x

18	478	8.4	660	10 BG473458	BG473458 602516255
19	474	8.3	694	10 BG910244	BG910244 602805735
20	463	8.1	463	9 A1356239	A1356239 qy65f11.x
21	462	8.0	462	9 A1309934	A1309934 q071e02.x
22	457	8.0	476	9 AA291351	AA291351 zt38e06.r
23	447	7.9	966	10 BG260547	BG260547 602372096
24	447	7.9	1063	10 BM462581	BM462581 AGENCOURT
25	445	7.8	517	10 BE881579	BE881579 QV3-ET019
26	441	7.8	987	10 BG170054	BG170054 602321514
27	436	7.7	436	9 A1192867	A1192867 qe68b05.x
28	436	7.7	442	9 A1990048	A1990048 wsl9e01.x
29	428	7.5	600	10 BE299145	BE299145 601118731
30	427	7.5	427	9 AA293275	AA293275 zt38e06.s
31	426	7.5	429	9 A136967	A136967 z102906.r
32	426	7.4	477	10 BG283908	BG283908 602407659
33	422	7.4	452	10 BF508040	BF508040 UI-H-B14-
34	418	7.3	539	9 A1146646	A1146646 qb92h04.x
35	417	7.3	830	10 B1464081	B1464081 603302858
36	409	7.2	1040	10 BG745723	BG745723 602723832
37	403	7.1	446	9 A1369277	A1369277 qy91b11.x
38	396	7.0	560	10 B1861416	B1861416 603386632
39	395	6.9	403	9 A1865808	A1865808 wk86c06.x
40	395	6.9	545	9 A1129988	A1129988 qc50c03.x
41	394	6.9	394	9 A1241357	A1241357 qk17e05.x
42	391	6.9	497	9 AA307843	AA307843 EST178697
43	388	6.8	569	10 BF061374	BF061374 7j51f11.x
44	387	6.8	431	9 AA313349	AA313349 EST185260
45	367	6.5	540	9 AM242451	AM242451 xm99a04.x

ALIGNMENTS

RESULT 1
LOCUS B1917860 767 bp mRNA linear EST 16-OCT-2001
DEFINITION 603183990F1 NIH_MGC_121 Homo sapiens CDNA clone IMAGE:5247718 5',
mRNA sequence.
ACCESSION B1917860
VERSION B1917860.1 GI:16199788
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS 1 (bases 1 to 767)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11625 row: d column: 23
High quality sequence stop: 759.
Location/Qualifiers
1..767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5247718"
/clone_11b="NIH_MGC_121"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range

FEATURES

source

Page 2

[illegible]

```
Db 421 ACAAGATTAATTAAGAACCAAGAAAGGAGTCTGAAGGGCTTGGAGACATGTTCCAGT 480
Oy 3077 ttgcaaacatcagaaagatgcaagattgagaaacggtgtaaaataaatacaggaat 3136
Db 481 TTGGCAACATCGAAGAAAGATGACAGATTGACAAAACGGGTAAATTAATTAACAGAAAT 540
Oy 3137 ccttaacatcagaagagagagatacgaatgaagcagagagagagagagatcaagcca 3196
Db 541 CTTTACATCAAGAAAGAGAGAGATACGAAATGACAGCAGAGAGAGAGAGATTTCAAGCCA 600
Oy 3197 aaaccggaatttaggagagagagagagagagagagagagagagagagagagagagat 3256
Db 601 AACTCGAAGATTTAGGAGACGACAGAGCTCGAGAGCGTACTATGTAATTCAGATT 660
Oy 3257 ttcatcggacattgg 3272
Db 661 TTCATCGAGCATTTGG 676

RESULT 3
BM466811 1150 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT 6431318 NIH_MGC_67 Homo sapiens CDNA clone IMAGE:5503072
DEFINITION 5', mRNA sequence.
ACCESSION BM466811
VERSION BM466811.1 GI:18515853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1150)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM12142 row: d column: 17
High quality sequence start: 9
High quality sequence stop: 652.
FEATURES
location/Qualifiers
SOURCE 1..1150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5503072"
/clone_1lb="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 294 a 310 c 309 g 235 t 2 others
ORIGIN
Query Match 10.4%; Score 589; DB 10; Length 1150;
Best Local Similarity 99.8%; Pred. No. 1.2e-268;
Matches 639; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 3333 cgttagactcagagccagagagagagagagagatgattgttattgcccagatcaa 3392
Db 38 CGCTAGACCTCAGACCCACGAGGAGCATATGATGATCTGTATGCTCAAGTCAA 97
Oy 3393 gaagccgaggaattccaaacctcctgttagacagtaacgatacctcaggaaatca 3452
|||||
|||||

Db 98 GAAGCCGGGAATTCGAAACCCCTACCTGTAGACAGTAACAGATCAACTCTAGCAATCA 157
Oy 3453 tgatcgatcacagcgtctgaggcaagaatttcagcaagcaagaagagatgaagttaga 3512
Db 158 TGATCGGATACACCGCTCTTAGGCAAGAAATTTTCGCAAGCAAGCAAGATGAAGATGTA 217
Oy 3513 agatcgtcggcggacctaaagattttgaagcaaccttggccgaacgacgagccgagcaga 3572
Db 218 AGATCGTGGCGGAGCTATAGTTTGTAGCAACCCCTGGCCGAACGACGCGCGGCGACGCA 277
Oy 3573 gaggcgagcagcactcgtgttcgttgaggatgcaagagcgagcagagagagagcg 3632
Db 278 GAGCGGGCCACACTCGGTCTCGTGGAGGTGCAAGATGCAAGCGGACGCGCAGAGAGAGG 337
Oy 3633 cgagagctccagcagagcccaagcgccagcagcagctctcgtcctcggcaagcaggaana 3692
Db 338 CGAGAGCTCCGACAGAGCCGAGGCGCAGATGACAGCTCTCTCGGCAAGCAGGAAAAA 397
Oy 3693 tgcagctcgtctcccaaggaactcttgaggagcagaactactcccttggggaaggttcca 3752
Db 398 TCCAGACTCGGTCTCCAGAGACTTTGGAGAGCAGAACTACTCCCTGGGGAAAGCTTTCCA 457
Oy 3753 gagtgcgaagagaaacccaggtactccagctaccagagctccaggaagagctactgag 3812
Db 458 GAGTGCCAAAGAGAACCCAGGTACTCCAGTACCAAGGCTCCAGAGACGGCTACCTGGG 517
Oy 3813 aggacatgctccaagcagaggtcagtctggaactcagaagagctccttcgccaaggaaca 3872
Db 518 AGGACATGGCTTCACAGCCAGGAGGTCTGTGAAACTCAGAGACTCTTCGCCAGAGAAACA 577
Oy 3873 gaggcggaagagcagcagatgaagaagcagctccttcggaaggagccagcaatata 3932
Db 578 GAGCGCGAAGAGACAGACAGATGAAAGACAGCCTCTTCGAGGGGCCACAGAACATATGA 637
Oy 3933 ctctgataagaagttcagagaccacagtgtaagccctccc 3972
Db 638 CTCGTATTAAGAAAGTCCAGACCCGATTACGCCCTCTCC 677

RESULT 4
BG743714 870 bp mRNA linear EST 15-MAY-2001
LOCUS BG743714
DEFINITION 602632895F1 NCI_CGAP_Skn3 Homo sapiens CDNA clone IMAGE:4778302 5',
mRNA sequence.
ACCESSION BG743714
VERSION BG743714.1 GI:14054367
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 870)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10633 row: m column: 23
High quality sequence start: 5
High quality sequence stop: 844.
FEATURES
location/Qualifiers
SOURCE 1..870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4778302"
/clone_1lb="NCI_CGAP_Skn3"
```

Site_2: SalI: Cloned unidirectionally. Site_1: NotI:
Average insert size 1.5kb. Library primer: Oligo dT
Technologies. Note: This is a MCI-CGAP Library."
208 c 213 g 243 t

QY 4370 gctgattt
Matches 582; Conservative 0; Mismatches 265;
Local Similarity 10.28; Score 582; DB 10; Length 870;

CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arranged by: The I.M.A.G.E. Consortium
Clone Distributor by: Incyte Genomics, Inc.
Found through the I.M.A.G.E. Consortium (ILN1)
http://image.jil.gov
Plate: ILN1012 row: e column: 16
High quality sequence stop: 710
Location/Qualifiers
1..713

Top: 110.
Location/Qualifiers
1. 713

homo sapiens
/db_xref="taxon:9606"

```
clone_lib="NIH MGC 80M
```

note="Organism" (phage

Insertion: pcMV-SPOR

Note: this is a NIH MCO

12/17

best Local Similarity	9.9%	Score 563;	DB 10;	Length 713;
Matches 683; Conservative		Pred. No. 3.3e-256;		
0; Mismatches		0;		

gaps

gatacgaatgtagcgaqgaggaat
caggaatcctttacatcagagagc

[illegible]

120 CCGAATTTAGGGAACG 120

3279 tgaatcctgacgatgcgtgagtcggcgccgctgtga 3278

181 TGATGAGTTAACTGCTTCTCTATGTAAGTTCATCA
180

5339 acctcagagcccaaacaa
5356 -cttctatgaaggttccatgctctcaaa

...ACCTCAGAGCCACAGAGAGAGCGG...
 ...GAGCCTGtatgccaagtcaaa...

...GAGATGCCCAAGTCAAGAAGCC 390

3459
3458

360 GATACACCAATTCAGCAAGCAAAACCAACT-
359

3519 tcggcggaacgtatata
cagcaggcaagaacgagtagaacatttgcg 3518

[illegible]

5'-gacactcgtgtccgtggaatgaa-3'
|||||
3'-acacacggccggacgcagagcc-5'

[illegible]

540 cttccagcgcagtcacagctctcttgcctc
539cagcgcgagag

3699 ctccatctca-
3698 -acgacag

600 CTCGGTCTCCACCAcctcctgqqaanctt

.....GAGGAGGCTTCCACCTGCTT

RESULT	5	Accession	713 bp	mRNA	linear	EST	06-FEB-2001
BI0170486		60232859F1					
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
		human					
		Homo sapiens					
		Eukaryota;					
		Mammalia;					
		1 (bases					
		NIH-MGC					
		http://mgc.nci.nih.gov/					
		National Institutes of Health,					
		Unpublished (1999)					
		Contact: Robert Strausberg,					
		Email: grants.rem@nih.gov					
		Tissue Procurement: ATCC					

3399 gcggatattcccaaacctcaactcaactgaagaagtaaaagatacaactcagaatcatgatgc 3398
300 gcggatattcccaaacctcaactcaactgaagaagtaaaagatacaactcagaatcatgatgc 299
3439 gatacagctctggaggcaagaattccagcaagcaagaagaatgaatattgaaatcgc 3458
360 gaaacacgctctgaagcgaagaatttcagcaacgaacaaacgaatgaatattgaaatcgc 359
3519 tcggaggacctaatagtttgaagcaacctggcccaaaagcaagccggcgaaagcgaaagcg 3518
420 tcggagcaacctaatagtttgaagcaacctggcccaaaagcaagccggcgaaagcgaaagcg 419
3579 gggcaactcgggtctccctgaaagtgtaaaatcaacgcagcagcgcggaaggaagcgaaag 3578
480 gcaacactctgggtctccctgaaagtgtaaaatcaacgcagcagcggcggaaggaagcgaaag 479
3639 ctcccaaggaagcccaagcgccagttcaagctcttcgacttcgaaagcaggaataatggc 3638
540 ctcccaaggaagcccaagcgccagttcaagctcttcgacttcgaaagcaggaataatggc 539

Oy	3759	caaaagagaaccacgactctccag	3782
Dd	660	CMAAGAGAACCACGATCTCCAG	683
RESULT	6	A1079538	536 bp mRNA linear EST-29-SEP-1998
Locus	A1079538/c		
DEFINITION	clone IMAGE:1674370 3', mRNA sequence.		
Accession	A1079538		
VERSION	A1079538.1	GI:3415789	
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
TITLE	Unpublished (1997)		
JOURNAL	Contact: Robert Strausberg, Ph.D.		
COMMENT	Email: cgapdb-remail.nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1447 Std Error: 0.00 Seq primer: -40ml3 fwd, ET from Amersham High quality sequence stop: 445. Location/Qualifiers 1..536 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="1674370" /clone_lip="Soares_fetal_liver_spleen_INFLS_S1" /sex="male" /dev_stage=""20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pTR3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI, this is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oliga(dT) primer [5' AACGGGAAGATTAAATAAAGACTGTCTTTTTTTTTTTTTTTTTT 3']," double-stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested wltah Pac I and cloned into the Pac I and Eco RI sites of the modified pTR3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	174 a	126 c	114 g 122 t
ORIGIN			
Query Match	9.4%; Score 536;	DB 9;	Length 536;
Best Local Similarity	100.0%;	Pred. No. 2.ee-243;	Mismatches 0;
Matches 536;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;		
OY	5144	aggcggtcacatcatcacaagaagaaccttttcggggtcccacagcggtggtgcgaagaagt	5203
Dd	536	AGCGCCTTCACATCACACAGAGAACAATTTTGCGGCCACACACGCCTGTCGACGAGAAGCT	477
OY	5204	cactctcgctcgaagattagagtggtgtgtgtgtgtcctcggggagctcgttgtctccaac	5263
Dd	476	CACITCTCGCGTCAGATTAAGAAGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	417
OY	5264	tctcttcatgttctgaacaactctgttatgttaaacatgatgtgggtgtctaagtgtct	5323
Dd	416	TTCCCTTCATTGTTTCGAACAATCTGTATTGTAAACCAAGCATGCGTGCTTAAGTCCCG	357
OY	5324	tgaattccccgtagtgaanaaatcttggaagtggaagctcagcatacatgtattacttta	5383
Dd	356	TGAATCCCAGATGTGAAAACCTGGAGGTGAAGCTCAGCATACCATGATATTACTGTAA	297

OY	5384	aaacgaaaaaaaagcagatgatctgtcatgttcatttttcttatggcaaatgtatctt	5443
Db	296	AAACGAAAAAAGACATGTAATGGTATGCTATTATTTTATTTATGGCACAATTGTATT	237
OY	5444	tctgttgacttgtttttaagaatatgtgtgcacacacagtaaccggtgtctcttcgat	5503
Db	236	TTCGTGTGACTGTGTTTTAGAATAATGATGTGNCACACAGTAACCGTGTCCTTGCAAT	177
OY	5504	tctgtgtcagtgtgtctgttcttctaatacagtgcgagggtgtctaaagtgttaccaagt	5563
Db	176	TTCGTGTGACATGCTCTCTTTCTTATATACAGCGCGCGGTCTGAAGTGAGGTTCACAGT	117
OY	5564	gtacgcgagatggaaccttgatgacagatgagctcttcttcacaagccccccttagctgtgag	5623
Db	116	GTAGCGCAGATGACCTTGGATGACAGTGGCTTTTCTCACAGCTCCCTCGAGCTGTGAG	57
OY	5624	aaacagcttctctgttacatalgtcaactctcataataaaggacataattctctgtt	5679
Db	56	AAACGCTTCTCTCTGTACATATGCACATCCCTAATAAAGCATATTCTCTCTGTT	1
RESULT	7		
LOCUS	A1497808/c	536 bp	mRNA linear EST 14-APR-1999
DEFINITION	tm9ff02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2165307 3',		
ACCESSION	A1497808		
VERSION	A1497808.1	GI:4389790	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.		
TITLE	1 (bases 1 to 536)		
	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
	National Cancer Institute / National Institute of Neurological		
	Disorders and Stroke, Brain Tumor Genome Anatomy Project		
	(CGAP/BTCGP), Tumor Gene Index		
JOURNAL	Unpublished (1998)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-femail.nhl.gov Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfield M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILMW at: www-bio.llnl.gov/bdrrp/image/image.html Insert Length: 779 Std Error: 0.00 Seq primer: -40UP from GlDco High quality sequence stop: 468. Location/Qualifiers 1..536 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2165307" /clone_id="NCI.CGAP.Brn25" /tissue_type="anaplastic oligodendroglioma" /lab_host="DH10B" /note="Organ: brain; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTATCAATCTGAAGTGGAGCGGCCGCAATGAGTGTATTTTATTTTATTTTATTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	174 a	126 c	114 g 122 t
ORIGIN			

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-femail.nih.gov
The sequence contained an oligo-dr track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrrp/image/image.html
Seq primer: M13 forward
POLA-Yes.

FEATURES
source

Location/Qualifiers
1. 608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3068738"
/lab_host="NCI-CGAP_Subs5"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; NCI-CGAP_Subs5
is a subtracted library derived from NCI-CGAP_Sub4. The
NCI-CGAP_Subs5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI-CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI-CGAP_K1d3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Cloneids
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI-CGAP_K1d5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Cloneids 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI-CGAP_Lus pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Cloneids 1414920-1417991, 1520904-1522439
); NCI-CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
3733-3735 (IMAGE Cloneids 1257096-1258631, 1469654-1470983,
1475592-1476743); NCI-CGAP_P12 pool 1 LLAM 2457-2459,
2758-2759, 3062-3068 (IMAGE Cloneids 985608-986759
, 1101192-1101959, 1217928-1220615); NCI-CGAP_Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Cloneids 1057416-1061255
, 114584-1145351). (10% of the driver population), plus a
pool of 3,840 arrayed clones from NCI-CGAP_Sub1 (IMAGE
Cloneids 2708616-2710535) and NCI-CGAP_Sub2 (IMAGE
Cloneids 2710536-2712455) (10% of the driver population
, plus a pool of 11,136 clones from NCI-CGAP_Sub3 (IMAGE
Cloneids 2712456-2723591) (10% of the driver population),
plus a pool of 5,472 clones from NCI-CGAP_Sub4 (IMAGE
Cloneids 2723592-2728969) (70% of the driver population).
Subtraction was performed as previously described [Bonaldi
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches to Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG_LIB=NCI-CGAP_Brn26
TAG_TISSUE=brain
TAG_SEQ=ATAGC

BASE COUNT 131 a 154 c 115 g 208 t

Query Match 9.3%: Score 530; DB 9; Length 608;

Best Local Similarity 99.8%: Pred. No. 1,76-240; Mismatches 1; Indels 0; Gaps 0;

Matches 580; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

2411 ccgactgctcttgaagcagatgtgacagcttctgtcttccaagagaagattg 2470
|||||
594 CCGACTGCTCTTTAGTGCATGTGATTCAGTTCTTTCACAGAGAAGATTG 535
|||||
2471 gaagtcagatgctcagaaacagacaagcaatttcagatgcagtcgaattgat 2530
|||||
534 GACGTCAAGATGTCAGAAAAACGACAAAGCAATTTTCAGATGCCAGTCAATGTGATT 475
|||||
2531 tcgctaaacacgaaatacaaaagcagatgattagatgctgacagagactaacca 2590
|||||
474 TCGTTAAACACGAAATCAAAAGCATGATTAGGTATAGCTGACGAGACTAACTCA 415

QY 2591 atacagtgatgaccagaagaagcaggtctccacagacagatgtygtccctccggtc 2650
|||||
DB 414 ATACAGTGGATGACCAAGAAAGCAGGTCTCCACAGAGAGATGTGGTCTTCCCTGGGTC 355
|||||
QY 2651 tgaagaatcaagctcgttgtagagatctgcagaccgagtgccagtgacttgatg 2710
|||||
DB 354 TGAAAGATGCAAGCTCGTTGGAGCTGACACCGCAGTGGCCAGGTGACCTTTGAATG 295
|||||
QY 2711 gggatctccttcgatctgctccagcccgagatatacagaagcgagtgatccaatg 2770
|||||
DB 294 GGGATTCCTTCCATGCTCCACGCCCGGATATATCAGAGCGAGGAGTCAATGAGA 235
|||||
QY 2771 gcttcagagctgcacatcttaataaaccgcggtagatgataatgaa 2830
|||||
DB 234 GCTTCAGAGCTGCCATCGACAAATCTTNGATNAACCCGTGTATATGATGTAAG 175
|||||
QY 2831 gcatgagaccttggaagaagacacagaagaagttaagatcaggagagagctgtat 2890
|||||
DB 174 GCATGAGACCTTGGAAAGACACAGAAAGTTCAAAGATCAGAGGAGAGCTGTAT 115
|||||
QY 2891 ccagagcagatgacgctccacctctctgagagacaaatgaaatgaaacaaaga 2950
|||||
DB 114 CCACGCCAGTATCAGCTTCCTCCTCTGAGAGACAAATGGAATGAAACCAAGAGA 55
|||||

QY 2951 aagtgataagactgatagaagaagaagataaactgaaa 2991
|||||
DB 54 AAGTGATAGACTGATGAAAAAAGATTAACCTGAAAA 14
|||||

RESULT 10
AL59694
LOCUS
DEFINITION DKF2p313f1730.r1 313 (synonym: h1cc2) Homo sapiens cDNA clone
AL59694
ACCESSION AL59694
VERSION AL59694.1 GI:15162982
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Koehler,K., Beyer,A., Mewes,W., Well,B. and Wiemann,S.
TITLE EST (Koehler,K., Beyer,A., Mewes,H.W., Well,B. and Wiemann,S.)
JOURNML Unpublished (1999)
COMMENT Contact: Koehler K

MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
No sl sequence available.

This clone (DKF2p313f1730) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcentrum Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source

Location/Qualifiers
1. 592
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p313f1730"
/clone_lib="313 (synonym: h1cc2)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pT73D-Pac; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

BASE COUNT 187 a 141 c 133 g 128 t 3 others

REFERENCE	AUTHORS
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo. 1 (bases 1 to 645)	Dias Neto,E., Garcia Correia,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zaio,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G.

FEATURES	SOURCE
Location/Qualifiers	1. 645
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone_id="PR0387"	
/dev_stage="Adult"	
/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESETES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	
BASE COUNT	122 a 176 c 115 g 232 t
ORIGIN	

Oy	2870	gaacagggaaggaagcctctatcccaagccagatgatacaacctcccaactctcgaggaagac	2929
Db	252	GATCAGGGAGAGAGCTCTGTATTCACAGCCAGTGAATCAGCCTTCCACTCTCTGGAGAGAC	193
Oy	2930	aattgaatcggaaaccaagagagaaggttgaataagactgataaaaaaaggataaaactggaa	2989
Db	192	AAATGAATGGAAACCAAGAAAGGTGATTAAGACTGATTAAGAAAAAAGGATTAATACTGGA	133
Oy	2990	aagaaagaagaagaatctagagataaaggagaagataaaatgaagaagccaaagaaggaaatgc	3049
Db	132	AAGAAAGAACAAATATGAGATTAAGGAGAAAGATTAATGAATAACCAAGAAAGCAATGCC	73
Oy	3050	tgaagggcttggagagacatgttcacggt	3077
Db	72	TGAAAGGCTTGGAGACATGTTTCAGGTT	45

RESULT	13
AM780417/c	
LOCUS	599 bp mRNA linear EST 12-MAY-2000
DEFINITION	ho21bo08.x1 NCI CGAP Co14 Homo sapiens cDNA clone IMAGE:303807 3' similar to TR:Q92340 Q92340 ATYPICAL PKC SPECIFIC BINDING PROTEIN.
	;', mRNA sequence.
ACCESSION	AM780417
VERSION	EST.
KEYWORDS	GI:7795020
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1. (bases 1 to 599)
AUTHORS	Moskaluk, Eugene D.; Finkbeiner, Christopher; Hootenhouse, Ronald.
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT	Tumor Gene Index Unpublished (1997)
	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CGNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.lnl.gov Possible reversed clone: polyA not found Seq primer: -40up from Gibco High quality sequence stop: 412. Location/Qualifiers 1..599
FEATURES	
SOURCE	

Oy	2750	gaggcaggggagtgtcatgtagaagcttcagaagctgcacatcgacaacattatgataaacccg	2809
Db	372	gagcagggcgatgcgaatagaccttcaagagctgccatgcaccaaatcttatgtataaacgg	313
Oy	2810	cggtaagatgatgatgatatgaagagccttggaagaagacacagaagaagttcaa	2869
Db	312	cggtagatgaatgatcatgaagcacatgagaccttgaaagaaacacagaaagaaagttcaa	253

File: image.jpg
Plate: LLA012030 row: 9 column: 12

Query Match	8.6%	Score 492	DB 10	Length 809
Best Local Similarity	100.0%	Pred. No. 1.8e-222		
Matches 492	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1728	tggtcttaacctgattgycacacaggaagattctgcgatttaagcccaattatgaltc	1787		
DB 56	TCGTCCTTACACCTGATGGCCACACAGGAAATTCGACATTGAAAGCCCACTTATGATTC	115		
QY 1788	aggatctgcagcctctgctgcagtgtcaaaaggtaccggtcaaaagagacacacaga	1847		
DB 116	AGGATCTGACAGCCCTTGCTGTGAGTGTCAAGGTACCCGGTCAAAAGAGAACACAGCAGNA	175		
QY 1848	tttgggaatcttgcgaatccattattaaatggagagacagcatctaaagatggagagct	1907		
DB 176	TTTTGGGAATCTTTGTGCAATCCATTATTTAAAGAGAGAGCAGCATCTAAAGATGGAGAGCT	235		
QY 1908	tcgggtgaatagatccaatgatagacagtaaaatggagaaatccctgttggcaagacaaaca	1967		
DB 236	TCGGGTGAATAGTCAACTGATAGCAGTAAATGGAGAAATCCCTGTGGGCAAGACAAACCA	295		
QY 1968	aggtgcacatggaaacccctaagaagtctatgtcttactgaaggcaataagaggaatgat	2027		
DB 296	AGATGCCATGGAAACCTTAACAAGCTTATGCTACTGTGAAGGCAATTAACAGAGCAATGAT	355		
QY 2028	ccagctatcttgcgaaggagaaatgaagcaagtgcgaatgagctggaagtcaccttggagccc	2087		
DB 356	CCAGCTTATGCTTCCACAGAGAAATTAAGCAATGAGCTGAAGTCACTCTGGAGCCCC	415		
QY 2088	ccctggagcagcagctgcaccttgtaaaacagcgtttgatgatagagaacgaagaattccca	2147		
DB 416	CCCTGGACCTGACCTGCCATTGAAACAAGCGTTGGATGATGAGAACGAAGAATTTCCCA	475		
QY 2148	ttcccttaacagtgaggtattgagggcttgcagtaatgcgccagacagaatggcgccttaag	2207		
DB 476	TTCCCTCTTACAGTGGGATTTGAGGGCTTGAATGATCCCCAGCAGAAATGCTGCTCTAGC	535		
QY 2208	taagataaagggg 2219			
DB 536	TAGGATTAATGCC 547			

RESULT 15	BI870563	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BI870563	60339404F1	NIH_MGC_90	Homo sapiens CDNA clone IMAGE:5404067 5'	575 bp	mrna	linear	EST 11-OCT-2001	
REFERENCE	Mammalia; Euteleostomi; Primates; Catarrhini; Hominoidea; Homo.							
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .							
TITLE	Unpublished (1999)							
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)							
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rsb@nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ http://imgc.nih.gov							



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 11:12:43 ; Search time 102.97 Seconds
(without alignments)
13571.030 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaagtgaccgtgtgctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table:

Gapco_NUC
Gapco 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents.NA: *
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PCtus.COMB.seq: *
6: /cgn2_6/ptodata/2/1na/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the total being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	0.4	1272	3 US-08-946-026-36	Sequence 36, Appl
2	22	0.4	1330	3 US-08-888-077A-26	Sequence 26, Appl
3	21	0.4	1774	4 US-09-215-252-16	Sequence 16, Appl
4	19	0.3	39	3 US-08-961-083-371	Sequence 371, App
5	19	0.3	196	3 US-08-961-083-145	Sequence 145, App
6	19	0.3	1027	4 US-09-303-524A-1	Sequence 1, Appl
7	19	0.3	1628	4 US-08-883-515-3	Sequence 3, Appl
8	19	0.3	5836	4 US-09-233-086-2	Sequence 2, Appl
9	18	0.3	565	1 US-08-469-427A-4	Sequence 4, Appl
10	18	0.3	565	2 US-08-609-443B-4	Sequence 4, Appl
11	18	0.3	561	2 US-08-569-063C-4	Sequence 4, Appl
12	18	0.3	591	2 US-08-469-427A-6	Sequence 6, Appl
13	18	0.3	591	2 US-08-609-443B-6	Sequence 6, Appl
14	18	0.3	591	2 US-08-569-063C-6	Sequence 6, Appl
15	18	0.3	624	2 US-08-609-443B-12	Sequence 12, Appl
16	18	0.3	624	2 US-08-569-063C-12	Sequence 12, Appl
17	18	0.3	886	1 US-08-469-427A-1	Sequence 1, Appl
18	18	0.3	886	2 US-08-609-443B-1	Sequence 1, Appl
19	18	0.3	886	2 US-08-569-063C-1	Sequence 1, Appl
20	18	0.3	966	2 US-08-766-738-2	Sequence 2, Appl
21	18	0.3	1229	2 US-08-975-316-85	Sequence 85, Appl
22	18	0.3	1293	3 US-08-722-184-7	Sequence 7, Appl
23	18	0.3	1293	4 US-09-043-937A-11	Sequence 11, Appl
24	18	0.3	1343	1 US-08-197-792-30	Sequence 30, Appl
25	18	0.3	1343	1 US-08-459-850-30	Sequence 30, Appl
26	18	0.3	1343	1 US-08-459-214-30	Sequence 30, Appl
27	18	0.3	2563	2 US-08-553-436A-7	Sequence 7, Appl

C	28	18	0.3	2900	2 US-09-027-337-9	Sequence 9, Appl
	29	18	0.3	3153	4 US-09-080-625-3	Sequence 3, Appl
	30	18	0.3	3247	4 US-08-718-388-4	Sequence 4, Appl
	31	18	0.3	3336	4 US-09-080-625-2	Sequence 2, Appl
	32	18	0.3	3694	4 US-09-080-625-5	Sequence 5, Appl
	33	18	0.3	3877	4 US-09-080-625-4	Sequence 4, Appl
	34	18	0.3	7824	4 US-08-718-388-6	Sequence 6, Appl
	35	18	0.3	16382	4 US-08-718-388-8	Sequence 8, Appl
C	36	17	0.3	300	4 US-09-439-313-240	Sequence 240, Appl
	37	17	0.3	523	4 US-08-896-164-13	Sequence 13, Appl
C	38	17	0.3	570	1 US-08-469-427A-10	Sequence 10, Appl
C	39	17	0.3	570	2 US-08-609-443B-10	Sequence 10, Appl
C	40	17	0.3	570	2 US-08-569-063C-10	Sequence 10, Appl
C	41	17	0.3	573	2 US-08-290-665A-141	Sequence 141, App
C	42	17	0.3	573	5 PCT-US95-10398-141	Sequence 141, App
C	43	17	0.3	771	4 US-09-230-637-8	Sequence 8, Appl
C	44	17	0.3	950	4 US-08-960-780-12	Sequence 12, Appl
C	45	17	0.3	950	4 US-09-073-898-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-946-026-36
; Sequence 36, Application US/08946026
; Patent No. 6034218
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Mitcham, Jennifer L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,026
; FILING DATE: 07-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-946-026-36

Query Match 0.4%; Score 22; DB 3; Length 1272;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3609 gtagcgagcgagcgagcgagcgag 3630
|||||

```

? CURRENT FILING DATE: 1998-12-18
? PRIOR APPLICATION NUMBER: US 08/618,651
? PRIOR FILING DATE: 1996-03-19
? NUMBER OF SEQ ID NOS: 42
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO. 16
? LENGTH: 1774
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (158)..(1291)
US-09-215-252-16

Query Match
Best Local Similarity 100.0%; Score 21; DB 4; Length 1774;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 547 aaacactgctgggaattccta 567
| | | | | | | | | | | | | | |
Db 871 aaacactgctgggaattccta 891
```

```

Query Match          0.3%; Score 19; DB 3; Length 39;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y      1359 caaaaatagcgagaag 1377
|||||

Db       871 aacactgctggagctcctaa 891
|||||

RESULT 4
US-08-961-083-371
; Sequence 371; Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; OPERATING SYSTEM: HP Vectra 486/33
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; TELECOMMUNICATIONS INFORMATION: PB340P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO.: 371:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-371

```

DB 21 CAAAAAATAGCGAAGAGG 39

RESULT 5

US-08-961-083-145
Sequence 145, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8512

TELEFAX: (301) 309-8504

INFORMATION FOR SEQ ID NO: 145:

SEQUENCE CHARACTERISTICS:

LENGTH: 196 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-083-145

Query Match 0.3%; Score 19; DB 3; Length 196;
Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1359 caaaaataagcgaagag 1377
|||||

DB 11 CAAAAAATAGCGAAGAGG 29

RESULT 6

US-09-303-524A-1
Sequence 1, Application US/09303524A

Patent No. 6238873

GENERAL INFORMATION:

APPLICANT: CHAMBERS, JONATHAN K.

APPLICANT: STEWART, BRIAN R.

APPLICANT: AMES, ROBERT S.

APPLICANT: SARAU, HENRY M.

APPLICANT: FOLEY, JIM

APPLICANT: ARNOLD, ANNE ROMANIC

TITLE OF INVENTION: METHODS OF SCREENING FOR AGONISTS AND

TITLE OF INVENTION: ANTAGONISTS OF THE INTERACTION BETWEEN THE HUMAN KIAA0001

FILE REFERENCE: GP50007

CURRENT APPLICATION NUMBER: US/09/303,524A

CURRENT FILING DATE: 1999-04-30

PRIOR APPLICATION NUMBER: 60/083,957

PRIOR FILING DATE: 1998-05-01

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 1027

TYPE: DNA

ORGANISM: Homo sapiens

US-09-303-524A-1

Query Match 0.3%; Score 19; DB 4; Length 1027;
Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 457 actctgtcagtagtagta 475
|||||

DB 420 actctgtcagtagtagta 438

RESULT 7

US-08-883-515-3/C
Sequence 3, Application US/08883515

Patent No. 5981836

GENERAL INFORMATION:

APPLICANT: Osteryoung, Katherine W

TITLE OF INVENTION: PLANT CHLOROPLAST DIVISION GENES

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pluckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/883,515

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27,386

REFERENCE/DOCKET NUMBER: 920905, 90016

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-9166

TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1628 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 3..1316

US-08-883-515-3

Query Match 0.3%; Score 19; DB 2; Length 1628;
Best Local Similarity 100.0%; Pred. No. 23;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 269 gccacagtcgcagttccac 287
|||||

DB 544 gccacagtcgcagttccac 526

RESULT 8

US-09-233-086-2
Sequence 2, Application US/09233086
Patent No. 6337192
GENERAL INFORMATION:
APPLICANT: Bartel, Paul L.
APPLICANT: Tavtigian, Sean V.
TITLE OF INVENTION: Myriad Genetics, Inc.
FILE REFERENCE: MMSCI - An MMAC1 Interacting Protein
CURRENT APPLICATION NUMBER: US/09/233,086
EARLIER FILING DATE: 1999-01-19
PRIOR APPLICATION NUMBER: US 60/071,861
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 5836
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (115)..(5757)
US-09-233-086-2

Query Match 0.3%; Score 19; DB 4; Length 5836;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1406 aaggttgagtcagcat 1424
DB 2195 aaggttgagtcagcat 2213
|||||

RESULT 9

US-08-469-427A-4/c
Sequence 4, Application US/08469427A
Patent No. 5607918
GENERAL INFORMATION:
APPLICANT: Eriksson, Ulf
APPLICANT: Olofsson, Birgitta
APPLICANT: Alltalo, Karl
APPLICANT: Pajusola, Katri
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
TITLE OF INVENTION: DNA CODING THEREFOR
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,427A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/397,651
FILING DATE: 01-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Evans, Joseph D
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 41979cp2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 565 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
TISSUE TYPE: adult mouse heart
US-08-469-427A-4

Query Match 0.3%; Score 18; DB 1; Length 565;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3608 tgcagcgagcgagcgag 3625
DB 482 tgcagcgagcgagcgag 465
|||||

RESULT 10

US-08-609-443B-4/c
Sequence 4, Application US/08609443B
Patent No. 5840693
GENERAL INFORMATION:
APPLICANT: ERIKSSON, Ulf
APPLICANT: OLOFSSON, Birgitta
APPLICANT: ALLTALO, Karl
APPLICANT: PAJUSOLA, Katri
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
TITLE OF INVENTION: DNA CODING THEREFOR
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,443B
FILING DATE: 01-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/397,651
FILING DATE: 01-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,427
FILING DATE: 06-JUN-1995
APPLICATION DATA:
APPLICATION NUMBER: US 08/569,063
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/41979CP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO

```

; ORIGINAL SOURCE:
; TISSUE TYPE: adult mouse heart
US-08-609-443B-4

```

Query Match	0.3%	Score 18:	DB 2:	Length 565:
Best Local Similarity	100.0%	Pred. No. 72:		
Matches 18:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0

Oy	3608	tcgacgcygcagcygcagc	362
Db	482	TGCAGCGGCAGCGGCAGG	465

RESULT 11
US-08-569-063C-4/c
; Sequence 4, Application US/08569063C
; Patent No. 5928939

Query Match	0.3%	Score 18;	DB 2;	Length 565;
Best Local Similarity	100.0%	Pred. No. 72;		
Matches 18; Conservative	0;	Mismatches	0;	Gaps 0;
07 3608	tgacggcagcgagcagcag	3625		

Db 482 TGCAGCGGCAGCGGCAGG 465

RESULT 12
US-08-469-427A-6/c
; Sequence 6, Application US/08469427A
; Patent No. 5607918

Query Match 0.3%; Score 18; DB 1; Length 591;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 18; Conservative 0; Indels 0; Gaps 0;

DY 3608 tgcagcgcgacgcgcagg 3625
 |||||
Db 508 TGCAGCGCAGCGCAGG 491

RESULT 13
 US-08-609-443B-6/c
 : Sequence 6, Application US/08609443B
 : Patent No. 5840693
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: ERIKSSON, Ulf
 :
 : APPLICANT: OLOFSSON, Birgitta
 :
 : APPLICANT: ALITALA, Karl
 :
 : APPLICANT: PAUSOLA, Kalri
 :
 : TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
 : TITLE OF INVENTION: DNA CODING THEREFOR

NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,443B
FILING DATE: 01-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/397,651
FILING DATE: 01-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,427
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569,063
FILING DATE: 06-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/41979CP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 591 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
TISSUE TYPE: adult mouse heart
US-08-609-443B-6

Query Match 0.3%; Score 18; DB 2; Length 591;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3608 tgcagcgagcgagcgagcg 3625
|||||
DB 508 tgcagcgagcgagcgagcg 491

RESULT 14
US-08-569-063C-6/c
Sequence 6, Application US/08569063C
Patent No. 5928939
GENERAL INFORMATION:
APPLICANT: ERIKSSON, Ulf
APPLICANT: OLOFSSON, Birgitta
APPLICANT: ALITALO, Karl
APPLICANT: PATUSOLA, Karl
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
COUNTRY: USA

ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,063C
FILING DATE: 06-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,427
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/397,651
FILING DATE: 01-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/41979CP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 591 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
TISSUE TYPE: adult mouse heart
US-08-569-063C-6

Query Match 0.3%; Score 18; DB 2; Length 591;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3608 tgcagcgagcgagcgagcg 3625
|||||
DB 508 tgcagcgagcgagcgagcg 491

RESULT 15
US-08-609-443B-12/c
Sequence 12, Application US/08609443B
Patent No. 5840693
GENERAL INFORMATION:
APPLICANT: ERIKSSON, Ulf
APPLICANT: OLOFSSON, Birgitta
APPLICANT: ALITALO, Karl
APPLICANT: PATUSOLA, Karl
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W., Suite 700
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,443B
FILING DATE: 01-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 11:13:58 ; Search time 547.74 Seconds

(without alignments)
17832.414 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaaagtacccgtgtgctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802:*

1: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2991	52.6	5510	22	AAH73337
2	1694	29.8	2718	22	AAH18587
3	1694	29.8	2718	22	AAH78369
4	1004	17.6	1800	22	AAH72793
5	842	14.8	867	22	AAH91875
6	725	12.7	823	22	AAH27320
7	725	12.7	823	22	AAH33594
8	407	7.2	556	22	AAH07109
9	395	6.9	425	22	AAH67583

10	269	4.7	597	22	AAH71480	Human cervical can
11	252	4.4	400	22	AAH66057	Novel human polynu
12	244	4.3	582	23	AAH68901	DNA encoding novel
13	234	4.1	540	21	AAH79294	Human lung tumour-
14	234	4.1	540	23	AAH23370	Human lung tumour-
15	211	3.7	510	23	AAH68902	DNA encoding novel
16	193	3.4	564	22	AAH12111	Human cDNA clone (
17	181	3.2	181	22	AAH74458	Human foetal liver
18	181	3.2	181	22	AAH54924	Probe #23610 used
19	181	3.2	556	22	AAH61963	Human foetal liver
20	181	3.2	556	22	AAH1889	Probe #10575 used
21	181	3.2	1262	23	AAH68903	DNA encoding novel
22	166	2.9	505	22	AAH61506	Human foetal liver
23	166	2.9	505	22	AAH29229	Human brain expres
24	166	2.9	505	22	AAH09806	Human bone marrow
25	166	2.9	505	22	AAH35700	Probe #7039 for ge
26	166	2.9	505	22	AAH17106	Probe #10101 used
27	166	2.9	505	22	AAH14115	Human cervical can
28	134	2.4	305	22	AAH72113	Human cervical can
29	134	2.4	305	22	AAH71048	Human cervical can
30	120	2.1	504	22	AAH70055	Human secreted exp
31	112	2.0	145	21	AAH45600	Human secreted exp
32	110	1.9	339	22	AAH69243	Human cervical can
33	71	1.2	180	21	AAH02894	Human secreted pro
34	52	0.9	2554	21	AAH01182	PD2 domain-compris
35	33	0.6	525	23	AAH67317	DNA encoding novel
36	33	0.6	525	23	AAH68814	DNA encoding novel
37	26	0.5	1090	20	AAH84303	Wheat geminivirus
38	23	0.4	508	22	AAH06145	Human cDNA clone (
39	23	0.4	700	22	AAH00331	Human reproductive
40	23	0.4	700	22	AAH62479	Human breast or ov
41	23	0.4	820	22	AAH14319	Human cDNA sequenc
42	23	0.4	875	22	AAH68551	Human protein HP10
43	23	0.4	127197	22	AAH61370	Soybean 515002 reg
44	22	0.4	396	23	AAH53141	Enterococcus faeca
45	22	0.4	1116	18	AAH78308	EGAP-binding prote

ALIGNMENTS

RESULT 1

AAH73337 standard; cDNA; 5510 BP.

AC AAH73337;

DT 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 4611.

KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

OS Homo sapiens.

PN WO200142467-A2.

PD 14-JUN-2001.

PF 08-DEC-2000; 2000MO-US33312.

PR 08-DEC-1999; 9905-0169681.

PR 21-DEC-1999; 9905-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210600.

PR 21-JUL-2000; 2000US-0220114.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Deeds J, Berger A, Zhao X;

DR WPL; 2001-375006/39.

sequence 5510 BP; 1526 A; 1242 C; 1418 G; 1305 T; 19 other;

	Mismatches	0;	Indels	1;	Gaps	1;
--	------------	----	--------	----	------	----

2358

[illegible]

2419 gttgcgagatgactttgaatcgccgcgataatc 2748

|||||||
agggcaggatgcattgagacttcgaagtgtccacgcacaatatcttatgatataaaccc 2808

Y 2809 gcgtagatgatgatgaagcgcacgtgaaccttgaagaacaacaaacttt

-3-ccccccc 2050

Y 2869 agatcaggaagaagantctctt-----
2598

2655 agaccaggagagagctctgtatccacagccagtgatcatcagcctccactctctggagaga 2658

2659 caatgtatgaaaccaagagaaatgataaactatatacaaaacacacgga 2988

[illegible]

3049 ctgaaggcgcttggagacatgttccaggtttgtgcaaacatcgaagaatgacacacatttc 3100

[illegible]

3160
3393ccaaacaaatacaggaatccttcatcagaagagagagatacgaatg 2898

2899 aaagcaggacgaggaattcaagccaaactcgagaatttagggatctgcacacgttacc 6070

2959 gagcgtgactatgctgaatttgcacgttgcacacccgctgctgatgatgagttta 3288

acgcacggggagttctctctatgaaggttcacatggtctcaacgctagacctcaagc 3348

3019 atgtatggggagattcttcttatgaaggttcacatgctgctctcaacgctagacctcagac 3078

[illegible]

OY 4429 atgacgtgcggccctgttctccctcaagttcctcttagctacatggatgacatcttagtc 4488
|||||
DB 4459 atgacgtgcggccctgttctccctcaagttcctcttagctacatggatgacatcttagtc 4218
|||||
OY 4489 tctgttgagagaaagtgaggagacatacacaataatggggcttctcgttaattcaag 4548
|||||
DB 4219 tctgttgagagaaagtgaggagacatacacaataatggggcttctcgttaattcaag 4278
|||||
OY 4549 cacgacatttctgtaacttctgtaactcactgttggtttctctctctctccagctctt 4608
|||||
DB 4279 cacgacatttctgtaacttctgtaactcactgttggtttctctctctctccagctctt 4338
|||||
OY 4609 gtgacggagagatcggtcatctctatcaagaagtagcacaatgttttgt 4668
|||||
DB 4339 gtgacggagagatcggtcatctctatcaagaagtagcacaatgttttgt 4398
|||||
OY 4669 caccatgggggtcctcttctgtaactgcttactcaacatcaataaagtgtatgc 4728
|||||
DB 4399 caccatgggggtcctcttctgtaactgcttactcaacatcaataaagtgtatgc 4458
|||||
OY 4729 ctggtctgcgttatacacaagctgttccgttccgttcacacgtgtgtgtacac 4788
|||||
DB 4459 ctggtctgcgttatacacaagctgttccgttccgttcacacgtgtgtgtacac 4518
|||||
OY 4789 cacagctctgtgtgaggaaagtgaggagacaggtggtcctcgtccggggagagctggcc 4848
|||||
DB 4519 cacagctctgtgtgaggaaagtgaggagacaggtggtcctcgtccggggagagctggcc 4578
|||||
OY 4849 tgcacattggccctgtgtctatcatgaggaggagctaaagaagaattctctctaggaa 4908
|||||
DB 4579 tgcacattggccctgtgtctatcatgaggaggagctaaagaagaattctctctaggaa 4638
|||||
OY 4909 gctcatggcccaacatcctcactaatattttaaattgatttctgtctgcacagctgtc 4968
|||||
DB 4639 gctcatggcccaacatcctcactaatattttaaattgatttctgtctgcacagctgtc 4698
|||||
OY 4699 aggaagggcacagaatgaggacagataaaccagatcatttgatctcctctca-c 5027
|||||
DB 4699 aggaagggcacagaatgaggacagataaaccagatcatttgatctcctctca-c 4758
|||||
OY 5028 gatttccaagtcagaggaattgattgtgttctaagtcatccacacctgtcagtctga 5087
|||||
DB 4759 gatttccaagtcagaggaattgattgtgttctaagtcatccacacctgtcagtctga 4818
|||||
OY 5088 ctggcgaatgcacagctcctcaatgattctgtcagtaaaacagaaagaaagcc 5147
|||||
DB 4819 ctggcgaatgcacagctcctcaatgattctgtcagtaaaacagaaagaaagcc 4878
|||||
OY 5148 gttacacatcacacagagaacatttcgggtcccaacagcgggtggtgcaggaagctcact 5207
|||||
DB 4879 gttacacatcacacagagaacatttcgggtcccaacagcgggtggtgcaggaagctcact 4938
|||||
OY 5208 ctgcgcgtcagatataaggtgtgtgtgtgtgtcgggagatctcgtgtcccatctcc 5267
|||||
DB 4939 ctgcgcgtcagatataaggtgtgtgtgtgtgtcgggagatctcgtgtcccatctcc 4998
|||||
OY 5268 tctcatctctgaaacatcctgtatgttaaacatggtcgtgggtctaaagtgcgtgtaa 5327
|||||
DB 4999 tctcatctctgaaacatcctgtatgttaaacatggtcgtgggtctaaagtgcgtgtaa 5058
|||||
OY 5328 tcccgatgtggaanaagctggaagtgaaagctcagcatcacatgatttactttaaac 5387
|||||
DB 5059 tcccgatgtggaanaagctggaagtgaaagctcagcatcacatgatttactttaaac 5118
|||||
OY 5388 agaaaaaagacatgtatgatatgtatatttttttatgtgcacatttatttgt 5447
|||||
DB 5119 agaaaaaagacatgtatgatatgtatatttttttatgtgcacatttatttgt 5178
|||||
OY 5448 gtgactgtttttagaataatgtgtccacacacgtacccgtgtctctctgcattct 5507
|||||
DB 5179 gtgactgtttttagaataatgtgtgtccacacacgtacccgtgtctctctgcattct 5238
|||||
OY 5508 gtgtcatgtgtctgttctcttaatacagtgcggggtgtcctaagtgtgttaccaggttac 5567
|||||

DB 5239 gtgtcatgtgtctgttcttctaatacagtgcggggtgtctaagtgtgttaccaggttac 5298
|||||
OY 5568 ggcacgtaaccttggatgacagtggtccttctcacaacccctccgtggtctggaac 5627
|||||
DB 5239 ggcacgtaaccttggatgacagtggtccttctcacaacccctccgtggtctggaac 5358
|||||
OY 5628 agcttctctgtacatgacacactcctaataaagacatattctctctgt 5679
|||||
DB 5359 agcttctctgtacatgacacactcctaataaagacatattctctctgt 5410
|||||
RESULT 2
AAH18587
ID AAH18587 standard; cDNA: 2718 BP.
XX
AC AAH18587;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18774.
XX
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isegaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 18774; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13631 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95883 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other;

Query Match 29.8%; Score 1694; DB 22; Length 2718;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1844; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 374 ttgggacacaccccttaagctctcgcgacaaatattccctctcaatgcttcgaacgagtaagt 433
Db 1 ttgggacacaccccttaagctctcgcgacaaatattccctctcaatgcttcgaacgagtaagt 433
OY 434 acccagctcaattggcctccctcactctctcgaatagatattcttcctgaagac 493
Db 61 acccagctcaattggcctccctcactctctcgaatagatattcttcctgaagac 493
OY 494 ctccaaggaataatcccaacgcctgctcaacaaacagctgctccctccaagcagaacactg 120
Db 121 ctccaaggaataatcccaacgcctgctcaacaaacagctgctccctccaagcagaacactg 120
OY 554 ctgggaagctcctaataacctgcgacaggaagaatgaataactacaagaagccctcccgagg 613
Db 181 ctgggaagctcctaataacctgcgacaggaagaatgaataactacaagaagccctcccgagg 613
OY 614 atactagtaactggtctcaacaaattcagaagacaaatgctcgtctctgagtgcca 673
Db 241 atactagtaactggtctcaacaaattcagaagacaaatgctcgtctctgagtgcca 673
OY 674 gtcaacccaatggtggcgaagtgctggaagaacaaacaggaatgaagatgaggaacagaag 733
Db 301 gtcaacccaatggtggcgaagtgctggaagaacaaacaggaatgaagatgaggaacagaag 733
OY 734 aggaataacagctgctgtaaacctctgtgacatgctgacacaggggttctggagcattaccca 793
Db 361 aggaataacagctgctgtaaacctctgtgacatgctgacacaggggttctggagcattaccca 793
OY 794 acttctctctggaatgattgttaagaactcgtaagaatgcccaacagatggaagggcctctg 853
Db 421 acttctctctggaatgattgttaagaactcgtaagaatgcccaacagatggaagggcctctg 853
OY 854 gaalccaatgtaagtccttcaagtcgcgaagcgcgacgaacccctggggttattagtaaac 913
Db 481 gaalccaatgtaagtccttcaagtcgcgaagcgcgacgaacccctggggttattagtaaac 913
OY 914 gattggagaaggtggtttaagctggaacatgtaaatcttctcgtggaatgattgattg 973
Db 541 gattggagaaggtggtttaagctggaacatgtaaatcttctcgtggaatgattgattg 973
OY 974 tcaggaattatgtagcgacacttcgaataagaatttgaacagaacacataatgattc 1033
Db 601 tcaggaattatgtagcgacacttcgaataagaatttgaacagaacacataatgattc 1033
OY 1034 gccaagcgaatgctgacacccatcttgggttccatggttcttcgacgaataaagaagc 1093
Db 661 gccaagcgaatgctgacacccatcttgggttccatggttcttcgacgaataaagaagc 1093
OY 1094 agtatagaacaactatcccaagtgagaagaacaaatattcaattcaagccggttttagccctg 1153
Db 721 agtatagaacaactatcccaagtgagaagaacaaatattcaattcaagccggttttagccctg 1153
OY 1154 acagcccgatattatggaacaacagagtggtgaacagtcgagggcttccaacgggtgacagag 1213
Db 781 acagcccgatattatggaacaacagagtggtgaacagtcgagggcttccaacgggtgacagag 1213
OY 1214 caacccgactgaacaccccgctgagcagatagactctcaacgactacccataagc 1273
Db 841 caacccgactgaacaccccgctgagcagatagactctcaacgactacccataagc 1273
OY 1274 caaccccgctgggaacacccacacatcgcctcgaagcctggaacctcgaatgattatgta 1333
Db 901 caaccccgctgggaacacccacacatcgcctcgaagcctcgaacctcgaatgattatgta 1333

```

```

OY 1334 cgactgtgaagcagtggttataacacaaataatgagcaagagcttaataatcagctta 1393
Db 961 cgactgtgaagcagtggttataacacaaataatgagcaagagcttaataatcagctta 1393
OY 1394 agaaaggtataagaaggtttgggattccagatcactccagagatgtataaatggtgct 1453
Db 1021 agaaaggtataagaaggtttgggattccagatcactccagagatgtataaatggtgct 1453
OY 1454 cagctccaatctatgtgaaataacatctcccccggggggcgcccatcgaatgagtcgcag 1513
Db 1081 cagctccaatctatgtgaaataacatctcccccggggggcgcccatcgaatgagtcgcag 1513
OY 1514 ttaagcaggaagacagacttaagagtgaaatgaglaattagttgggccaatcccaag 1573
Db 1141 ttaagcaggaagacagacttaagagtgaaatgaglaattagttgggccaatcccaag 1573
OY 1574 aggaagtggttctgctcttgaagaacacaaagatggaaggaactgtagcctctggtc 1633
Db 1201 aggaagtggttctgctcttgaagaacacaaagatggaaggaactgtagcctctggtc 1633
OY 1634 ttgcgcaggaagacgccttccacaaaggggaactgtaatgacagccaagccaatgacga 1693
Db 1261 ttgcgcaggaagacgccttccacaaaggggaactgtaatgacagccaagccaatgacga 1693
OY 1694 ttccaaagaagaagaagacagaagatgagatgttcttcaacccctgatatgacagga 1753
Db 1321 ttccaaagaagaagaagacagaagatgagatgttcttcaacccctgatatgacagga 1753
OY 1754 aatttctgacatttgaagttccacacttaattgaagatctgacagcccttggtgacag 1813
Db 1381 aatttctgacatttgaagttccacacttaattgaagatctgacagcccttggtgacag 1813
OY 1814 tcaaaagtaacccggtccaagaagaacacacagcagatttgggaactcttgtaagtcatta 1873
Db 1441 tcaaaagtaacccggtccaagaagaacacacagcagatttgggaactcttgtaagtcatta 1873
OY 1874 ttaatggagggagcgcacactcaaaagatggaaggtctcgggtgaatgaatgaatgaatga 1933
Db 1501 ttaatggagggagcgcacactcaaaagatggaaggtctcgggtgaatgaatgaatgaatga 1933
OY 1934 taaatggaagaatccctgttggcgaagaacaaacgaatgacatggaacacctaagaaggt 1993
Db 1561 taaatggaagaatccctgttggcgaagaacaaacgaatgacatggaacacctaagaaggt 1993
OY 1994 ctatgtcactgaagcgaataaaggaatgacaggaatgacagccttattgttgaagaagaataa 2053
Db 1621 ctatgtcactgaagcgaataaaggaatgacaggaatgacagccttattgttgaagaagaataa 2053
OY 2054 gcaaggtgcaatgagctgaagtcacacttggaagcccccctggaactgagctgacatgaa 1680
Db 1681 gcaaggtgcaatgagctgaagtcacacttggaagcccccctggaactgagctgacatgaa 1680
OY 2114 cagcgttgagatagagaacgaagaattcccatccctcctacagttggaatgagggc 2173
Db 1741 cagcgttgagatagagaacgaagaattcccatccctcctacagttggaatgagggc 2173
OY 2174 ttgattgattgcgccagcagaatgctgcccctcagtaggataatggt 2220
Db 1801 ttgattgattgcgccagcagaatgctgcccctcagtaggataatggt 2220

```

RESULT 3

AAH78369

ID AAH78369 standard; DNA; 2718 BP.

XX AAH78369;

XX 26-NOV-2001 (first entry)

XX Nucleotide sequence of a human protein kinase/protein phosphatase.

XX Human; protein kinase; protein phosphatase; signal transduction; ss.

OS	Homo sapiens.	Location/Qualifiers
XX		33..2630
FH	Key	CDS
FT		/*tag=a
FT		/product="protein kinase/protein phosphatase"
XX		
PN	MO200109316-A1.	
PD	08-FEB-2001.	
XX		
XX	28-JUL-2000; 2000WO-JP05061.	
XX		
PF	29-JUL-1999; 99JP-0248036.	
PR	18-OCT-1999; 99US-0159590.	
PR	11-JAN-2000; 2000JP-0118776.	
PR	17-FEB-2000; 2000US-0183322.	
PR	02-MAR-2000; 2000JP-0183767.	
PR	09-JUN-2000; 2000JP-0241899.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;	
PI	Senoo C, Nezu J;	
DR	WPI: 2001-570286/64.	
XX		
DR	P-PSDB; MAG67610.	
XX		
PT	New genes encoding proteins with protein kinase/protein phosphatase	
PT	activity, useful in the diagnosis and treatment of diseases -	
XX		
PS	Claim 1; Page 67-76; 233pp: Japanese.	
XX		
CC	The present sequence encodes a human protein kinase/protein phosphatase.	
CC	It is expected that the protein kinase/protein phosphatase gene	
CC	participates in signal transduction in cells. The protein kinase/protein	
CC	phosphatase polypeptides and polynucleotides are useful for developing	
CC	diagnostics and treatment agents for human and animal diseases. The	
CC	protein kinase/protein phosphatase polypeptides are useful as target	
CC	molecules in designing novel drugs. The protein kinase/protein	
CC	phosphatase polynucleotides are useful as a source of probes and	
CC	primers, which may be used to isolate homologous sequences.	
XX		
SQ	Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other;	
	Query Match 29.8%; Score 1694; DB 22; Length 2718;	
	Best Local Similarity 99.8%; Pred. No. 0;	
	Matches 1844; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
OY	374 ttgaagtcacacctcagctccttcgtgacgaataatccctcttaatgttgacgcagtagtg 433	
Db	1 ttgaagtcacacctcagctccttcgtgacgaataatccctcttaatgttgacgcagtagtg 60	
OY	434 accgcgcctaatgtgcctctccactctctgcagtgatagtaattttctctgaagac 493	
Db	61 accgcgcctaatgtgcctctccactctctgcagtgatagtaattttctctgaagac 120	
OY	494 ctccaagaaaatcccaacacgcgtgtcaacaacacagcttgcttctccaagcaaacatcy 553	
Db	121 ctccaagaaaatcccaacacgcgtgtcaacaacacagcttgcttctccaagcaaacatcy 180	
OY	554 ctgggaagtccttaaaacctcgcgcagcaagaagaatgaataactacagaagacccctccgcgg 613	
Db	181 ctgggaagtccttaaaacctcgcgcagcaagaagaatgaataactacagaagacccctccgcgg 240	
OY	614 ataccagtaaacggtcctaaccattccaagagacatcgcctcgtctctcgaagcca 673	
Db	241 ataccagtaaacggtcctaaccattccaagagacatcgcctcgtctctcgaagcca 300	
OY	674 gtccacccaatggttggtgcaagtgtgtagaagaacaagaatgaggaatggagacagaag 733	

Dh	301	g t c a c c c a a t g t g t g g c a a g t g g c a a g a g a a c a a g a c a g a t g s g a t t g g g a c a a g a g	360
Qy	734	a g a t a a c a g a t c g t t g t g a a c c t g t t g a a c t g t g a c a a g g t t t g g a g c a t a t a c c a	793
Dh	361	a g g a t a a c a g t c g t t g g a a c c t g t t g a a t c t g a c a c g g t t t g g g c a t a t a c c a	420
Qy	794	a c t t t c t c t g a t g a t a t a t g t a a a g c t g t a g a a g t c c c c a a c g a t g g a g g c t c t g	853
Dh	421	a c t t t c t c t g a t g a t a t a t g t a a a g c t g t a g a a g t c c c a a c g a t g g a g g c t c t g	480
Qy	854	g a a t c c a t g t a g t g c t t t c a g t g t c t a a g c g g c a g a a c c t g g g t a t a t a g t a a a a c	913
Dh	481	g a a t c c a t g t e g t c c t t c t a a t g t c t c g a a g c g g a c c c t g g g t t a t a t a g t a a a a c	540
Qy	914	g a t t g g a a a a a g g t g g t g a a a g c t b a a a c t g a a a a t c t t t t g t g a a t g g t t c a t g	973
Dh	541	g a t t g g a a a a a g g t g g t g a a a g c t g a a a c t g a a a a t c t t t t g t g a a a t g a t t g a t g	600
Qy	974	t c a g a t a t a t g a t g t g g a c c t t c g a a a t a t a g a a g a t t t g a c a a g c a c a a t a t g t t c	1033
Dh	601	t c a g a t a t a t a t g t g g a c c t t c g a a a t a t a g a a g a t t t g a c a a g c a c a a t a t g t t c	660
Qy	1034	g c c a a g c c a t g c t g a t a c a c c a t a t t g t t c a t g t g t c t g c a g a a a t a a a g a g c	1093
Dh	661	g c c a a g c c a t g c t g a t a c a c c a t a t t g t t c a t g t g t c t c t g a c a a a t a a a g a g c	720
Qy	1094	a g t t t g a a c a a c t a t c c c a a g t b a g a a g a a c a t t a c t a t t c a a g c g t t t a a g c g t	1155
Dh	721	a g t a t g a a c a a c t a t c c c a a g t b a g a a g a a c a a t a c t a t a t a a g c g t t t a a g c c g t	780
Qy	1154	a c a g c a g t a t a t g a c a a a g a g a t b t g a a a c a g t c a a g g t c t c a a c a g t g c a a a g a g	1215
Dh	781	a c a g c a g t a t a t g a c a a a c a g a g t g t g a a c a g t g a a g g c t c a a c a g t g c a a g a g	840
Qy	1214	c a c c c g a c t g a a c a c c c g c t g a g a g a t a g a t c a c t c a a g a t a c t c a t a g c	1277
Dh	841	c a c c c g a c t g a a a c a c c c g c t g a g a g a t a g a t c a c t c a a g a c t a c t c a t a g c	900
Qy	1274	c a c a c c c t c t g g a a a a c c a a c c g t c a c c t c a g c t c g a c c t c a g a a t g a t t a g t a	1333
Dh	901	c a c a c c c t c g g a a a a c c a c a c c g c t c a g c c t c g a c c t c a g a a t g a t t a t a g t a	960
Qy	1334	c g a c t g t a a g a g a t g t g t a t a a c c c a a a a a a a t a g c a a a g a g c t a a t a t c a a g t a	1395
Dh	961	c g a c t g t a a g a g a t g t g t a t a a c a c c a a a a a t a g c a a a g a g c t a a t a t c a a g t a	1024
Qy	1394	a g a a a g t a c a g a a g g t t t g g a t t c a g a c a c t c c a g a g a t g a c a a t a g t g t c t	1455
Dh	1021	a g a a a g t a c a g a a g g t t t g g a t t c a g a c a c t c c a g a g a t g a c a a t a g t g t c t	1080
Qy	1454	c a g t c c a a t a t a g t b a a a a a c t t c c c c c g g g g g g c a t t c a g t g a t g g c c a g c	1511
Dh	1081	c a g t c c a a t a t a g t b a a a a a c a c t t c c c c c g g g g g g c a t t c a g t g a t g g c c a g c	1144
Qy	1514	t t a a g c a g a g a c a g a c t t a t a g a g t a a a t g a g t a g a t t t a g t g g g c a a t c c c a a g	1577
Dh	1141	t t a a g c a g a g a c a g a c t a t a g a g t a a a t g a g t a g a g t t t a g t g g g c a a t c c c a a g	1200
Qy	1574	a g a a a g t g t t c g c t g t t g a g a a c c a a g a t g a a g a a c t g t a g c t t c t g t c t	1633
Dh	1201	a g a a a g t g t t c g c t g t t g a g a a c c a a g a t g a a g a a a c t g t a g c t t c t g t c t	1266
Qy	1634	t t c g c a a g a a g a g c t t c c a c c c a a g g a a c t g a a t g a a g c c a a g c c a g a t g c a a a	1693
Dh	1261	t t c g c a a g a a g a g c c t t c c a c c c a a g g a a c t g a a t g a a g c c a a g c c a a g t g c a a a	1321
Qy	1694	t t c c a a a g a a a c g a a a g c a a a g a t a g a g a t a t t g t c t t a a c c t g a t g g a c c a a g g	1755
Dh	1321	t t c c a a a g a a a c g a a a g a t a g a g a t a t t g t c t t a a c c t g a t g g a c c a a g g	1381
Qy	1754	a a t t c t g a c a t t t g a a g t c c c a c t t a a t g a t t c a a g a t c t g a a g c c t g t g t c a g t	1811
Dh	1381	a a t t c t g a c a t t t g a a g t c c c a c t t a a t g a t t c a a g a t c t g a a g c c t g t g t c a g t	1444

Db 545 tcgaatgtattagtagcagctgtaagcagctggtgtatcaacacaaatagcgcaag 604
 QY 1377 gcttaataatcagccttaagaaggtacagaaggtttgggaattcagcatcactcagaga 1436
 Db 605 gcttaataatcagccttaagaaggtacagaaggtttgggaattcagcatcactcagaga 1436
 QY 1437 tgaacaataagtgcttaagctccaatctatgtgaaacatctcccccggggggcgcc 664
 Db 665 tgaacaataagtgcttaagctccaatctatgtgaaacatctcccccggggggcgcc 664
 QY 1497 catcagaatgagcagcacttaagcagagacagacttatagagtgaaatgagtag 1552
 Db 725 catcagaatgagcagcacttaagcagagacagacttatagagtgaaatgagtag 1552
 RESULT 7
 ID AAH33594 standard; cDNA; 823 BP.
 AC AAH33594;
 XX
 DT 03-SEP-2001 (first entry)
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:650.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX colorectal carcinoma; ss.
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 PD 05-APR-2001.
 XX
 PE 28-SEP-2000; 2000WO-US26524.
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CN;
 DR WPI: 2001-235357/24.
 DR P-PSDB; AAG74163.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PR useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1; Page 2675; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens (P), where
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in gene
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the activity of P by expressing
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 XX Sequence 823 BP; 249 A; 183 C; 201 G; 188 T; 2 other;

Query Match 12.7%; Score 725; DB 22; Length 823;
 Best Local Similarity 99.98;
 Matches 775; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 777 tttagacatataaccacactttctctgatatgataagctcgtagaagctcccaa 836
 Db 5 tttagacatataaccacactttctctgatatgataagctcgtagaagctcccaa 836
 QY 837 cgtatgagggcctctgggaatccatgtatgctcttcaagtctcgaagcggaaccc 64
 Db 65 cgtatgagggcctctgggaatccatgtatgctcttcaagtctcgaagcggaaccc 64
 QY 897 ggggtatattagtaaaagcattgggaagaggtgtgtaaagctgaacataatcttttcg 124
 Db 125 ggggtatattagtaaaagcattgggaagaggtgtgtaaagctgaacataatcttttcg 956
 QY 957 tgaatgattgcatgttcaggaatataatgtagtcgaccttgaatataagatttgaca 184
 Db 185 tgaatgattgcatgttcaggaatataatgtagtcgaccttgaatataagatttgaca 184
 QY 1017 agcaacaataatgttcgcaagcagtcgtacaccacatcttggttcattgtgttc 244
 Db 245 agcaacaataatgttcgcaagcagtcgtacaccacatcttggttcattgtgttc 1076
 QY 1077 tgcagaataataaagcagtatgacaacactcccaaggtgaagaacaatactatc 304
 Db 305 tgcagaataataaagcagtatgacaacactcccaaggtgaagaacaatactatc 1136
 QY 1137 aagcgttttaagccttgacagcagtatattgacaacaggaagtgtgaacagtgcagc 364
 Db 365 aagcgttttaagccttgacagcagtatattgacaacaggaagtgtgaacagtgcagc 1196
 QY 1197 tcaacaggtgcagagagacaccccgactgaacacacccgctgcagagataagctcactc 424
 Db 425 tcaacaggtgcagagagacaccccgactgaacacacccgctgcagagataagctcactc 1256
 QY 1257 aagactactatagcgcacacccctcggaacacacacacacacacacacacacacac 484
 Db 485 aagactactatagcgcacacccctcggaacacacacacacacacacacacacacac 1316
 QY 1317 tcaaatgtatttagtagcagctgtagcaggtgtataacacacacacacacacacacac 544
 Db 545 tcaaatgtatttagtagcagctgtagcaggtgtataacacacacacacacacacacac 1376
 QY 1377 gcttaataatcagccttaagaaggtacagaaggtttgggaattcagcatcactcagaga 604
 Db 605 gcttaataatcagccttaagaaggtacagaaggtttgggaattcagcatcactcagaga 604
 QY 1437 tgaacaataagtgcttaagctccaatctatgtgaaacatctcccccggggggcgcc 664
 Db 665 tgaacaataagtgcttaagctccaatctatgtgaaacatctcccccggggggcgcc 1496
 QY 1497 catcagaatgagcagcacttaagcagagacagacttatagagtgaaatgagtag 1552
 Db 725 catcagaatgagcagcacttaagcagagacagacttatagagtgaaatgagtag 1552
 RESULT 8
 ID AAH07109 standard; cDNA; 556 BP.
 AC AAH07109;
 XX
 DT 26-JUN-2001 (first entry)
 DE Human cDNA clone (5'-primer) SEQ ID NO:3944.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.

XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 1; SEQ ID 3944; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 556 BP; 154 A; 129 C; 144 G; 126 T; 3 other;
SQ

Query Match 7.2%; Score 407; DB 22; Length 556;
Best Local Similarity 99.6%; Pred. No. 4,9e-185;
Matches 507; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 374 ttgaggtacacacctcagtccttcaggaacataatgccttcattcgttcgacgagtagtg 433
|||||
DB 1 ttgaggtcacaccttcagtccttcaggaacataatgccttcattcgttcgacgagtagtg 60
OY 434 accagccttaattggtccttcctcacttcgtcagtgatagtaattttctcttggaagac 493
|||||
DB 61 accagccttaattggtccttcctcacttcgtcagtgatagtaattttctcttggaagac 120
OY 494 cttaagaagaaatcccaacagctgtcaacaacagctgcttcctccaagcaagaacagc 553
|||||
DB 121 cttaagaagaaatcccaacagctgtcaacaacagctgcttcctccaagcaagaacagc 180
OY 554 ctggagagtccttaaaacctgacacaggaagaagatgaataactacagaagcctccgcggg 613
|||||
DB 181 ctggagagtccttaaaacctgacacaggaagaagatgaataactacagaagcctccgcggg 240
OY 614 atactagtagtgcgtctaaccaatttcagagagagaatctgcgtctcgttcgagtgcca 673
|||||

DB 241 atactagtagtgcgtctaaccaatttcagagagagaatctgcgtctcgttcgagtgcca 300
OY 674 gtcaaccaatgtggtggaagctggtgagaaacaaagacagatgagatgagacgaag 733
|||||
DB 301 gtcaaccaatgtggtggaagctggtgagaaacaaagacagatgagatgagacgaag 360
OY 734 aggatatacagctggttgaacctgttgacaatcgtcagacagggtttgagacataccca 793
|||||
DB 361 aggatatacagctggttgaacctgttgacatcgtcagacagggtttgagacataccca 420
OY 794 actttctcttgatgataatgataaagctcgttagaagtcctccaacagatgagggccttg 853
|||||
DB 421 actttctcttgatgataatgataaagctcgttagaagtcctccaacagatgagggccttg 480
OY 854 gaatccatgtagtgccttcagtgctcga 882
|||||
DB 481 gaatccatgtagtgccttcagtgctcga 509

RESULT 9
AAF67583
ID AAF67583 standard; cDNA; 425 BP.
XX
AC AAF67583;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 3345.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000MO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klingner J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Crain B;
XX
DR WPI: 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 1045; 1046pp; English.
XX

The present sequence is one of 3351 sequences in a library of human polynucleotides. The library is used to detect differentially expressed genes correlated with a cancerous state of a mammalian cell and can detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional copies of the polynucleotides. The probes can be used for chromosome mapping of the polynucleotide and for detection of transcription levels. Ribozymes or antisense oligonucleotides can be generated. The polynucleotides and their gene products are used as genetic or biochemical markers (e.g. in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or monitor the efficacy of therapies and preventive interventions. The polynucleotides, polypeptides and antibodies against them can be used in pharmaceutical compositions to treat the cancers and proliferative disorders such as neoplasia,

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN MO200175067-A2.
 XX 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HXSE-) HXSEQ INC.
 PA Drmanac RF, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 DR P-PSDB; ABG04715.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 1; SEQ ID No 4706; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (I) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 510 BP; 133 A; 133 C; 144 G; 100 T; 0 other;

Query Match 3.7%; Score 211; DB 23; Length 510;
 Best Local Similarity 99.2%; Pred. No. 8.7e-91;
 Matches 361; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3320 ccattgctcacaagcgtacagccacagagcagagggcatatgatgatcttgc 3379
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 145 ccattgctcacaagcgtacagccacagagcagagggcatatgatgatcttgc 204
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 3380 atgcccagtcagaagcgcggaattccaacccctcaacctgtagacagtaacagatcaa 3439
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 205 atgcccagtcagaagcgcggaattccaacccctcaacctgtagacacataacagatcaa 264
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 3440 ctccctagcaatcatgctgagatcacgctctgagcaagaattcagcaagcaagaag 3499
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 265 ctccctagcaatcatgctgagatcacgctctgagcaagaattcagcaagcaagaag 324
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 3500 atgaagatgtagaagatcgtcgcgcgacatagtttgagcaaccctgscgaacgcac 3559
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 325 atgaagatgtagaagatcgtcgcgcgacatagtttgagcaaccctgscgaacgcac 384

OY 3560 ggcgcgcagcagcagagcgcgcgcacactcgtctcgttgagagtgacagatgcagcgcagc 3619
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 385 ggcgcgcagcagcagatagcgcgcacactcgtctcgttgagagtgacagatgcagcgcagc 444
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 3620 ggcagagagcgcgcgcagcagcctccacagcagccacagcagcagcagcagcagcagc 3679
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 445 ggcagagagcgcgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 504
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 3680 aaag 3683
 ||||
 DB 505 aaag 508

Search completed: July 24, 2002, 14:29:07
 Job time: 11709 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 09:16:33 ; Search time 6663.36 Seconds
(without alignments)
17866.522 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689
Sequence: 1 atgaagtgaccgtgtgctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_lm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	4848	85.2	5958	9	AF196185	AF196185 Homo sapi
2	4020	70.7	4071	9	AF467002	AF467002 Homo sapi
3	3411	60.0	4062	9	AF467003	AF467003 Homo sapi
4	3024	53.2	3960	9	AF467004	AF467004 Homo sapi
5	2577	45.3	2696	9	AF000761	AF000761 Homo sapi
6	2414	42.4	3470	9	AF196186	AF196186 Homo sapi
7	2003	35.2	196374	2	AL591464	AL591464 Homo sapi
8	1952	34.3	106198	9	AL160409	AL160409 Human DNA
9	1952	34.3	138688	2	AC025823	AC025823 Homo sapi
10	1899	33.4	2116	9	AK024668	AK024668 Homo sapi
11	1762	31.0	22279	2	U82210	U82210 Homo sapien
12	1694	29.8	2718	6	BD004905	BD004905 Novel gen
13	1694	29.8	2718	9	AK027735	AK027735 Homo sapi
14	1670	29.4	2967	9	AF332592	AF332592 Homo sapi
15	1670	29.4	3822	9	AF332593	AF332593 Homo sapi
16	1670	29.4	3933	9	AF467006	AF467006 Homo sapi
17	1497	26.3	1592	9	BC011711	BC011711 Homo sapi
18	1494	26.3	2004	9	AF177228	AF177228 Homo sapi
19	1452	25.5	3801	9	AF252293	AF252293 Homo sapi
20	958	16.8	3735	9	AF467005	AF467005 Homo sapi
21	714	12.6	1772	9	AK025892	AK025892 Homo sapi
22	427	7.5	427	6	AX332939	AX332939 Sequence
23	395	6.9	425	6	AX072873	AX072873 Sequence
24	387	6.8	104357	9	AL392123	AL392123 Human DNA
25	269	4.7	597	6	AX187060	AX187060 Sequence
26	252	4.4	400	6	AX071341	AX071341 Sequence
27	245	4.3	165146	2	AC011644	AC011644 Homo sapi
28	245	4.3	167163	2	AC013705	AC013705 Homo sapi
29	245	4.3	169337	2	AL356462	AL356462 Homo sapi
30	245	4.3	193131	9	AL390766	AL390766 Human DNA
31	234	4.1	540	6	AX321805	AX321805 Sequence
32	181	3.2	109816	9	AL360233	AL360233 Human DNA
33	181	3.2	138322	2	AC012241	AC012241 Homo sapi
34	181	3.2	196259	2	AL392171	AL392171 Homo sapi
35	134	2.4	305	6	AX187694	AX187694 Sequence
36	134	2.4	309	6	AX186627	AX186627 Sequence
37	121	2.1	173987	9	AL450337	AL450337 Human DNA
38	120	2.1	504	6	AX185634	AX185634 Sequence
39	115	2.0	167163	2	AC013705	AC013705 Homo sapi
40	110	1.9	339	6	AX184822	AX184822 Sequence
41	104	1.8	176405	9	AL138768	AL138768 Human DNA
42	96	1.7	5500	10	AB005549	AB005549 Rattus no
43	91	1.6	4167	10	AY026057	AY026057 Mus muscu
44	37	0.7	85713	9	HSY237C10_3	Continuation (4 of
45	37	0.7	168071	2	AL391479	AL391479 Homo sapi

ALIGNMENTS

RESULT 1	AF196185	5958 bp	mRNA	linear	PRI 16-NOV-2001
LOCUS	AF196185		Homo sapiens atypical PKC isotype-specific interacting protein long		
DEFINITION	AF196185		Variant mRNA, complete cds.		
ACCESSION	AF196185		GI:13491609		
VERSION	AF196185.1		GI:13491609		
KEYWORDS					
SOURCE			human.		
ORGANISM			Homo sapiens		
			Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;		
			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE			1 (bases 1 to 5958)		
AUTHORS			Fang,C.M. and Xu,Y.H.		
TITLE			Down-regulated expression of atypical PKC-binding domain deleted		
			asip isoforms in human hepatocellular carcinomas		
JOURNAL			Cell Res. 11 (3), 223-229 (2001)		
MEDLINE			21520080		
PUBMED			11642408		
REFERENCE			2 (bases 1 to 5958)		
AUTHORS			Fang,C. and Xu,Y.		
TITLE			Exon/Intron Structure and Splicing Variants of a Novel Human		
			Polarity Gene, Hasip		

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5958)
AUTHORS Fang, C. and Xu, Y.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Laboratory of Molecular and Cellular
Oncology, Shanghai Institutes for Biological Sciences, Chinese
Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P.R. China

FEATURES
Location/Qualifiers
1. 5958

source

CDS

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
map="10p11.2"
288..4349
/note="Hs15p1a"
/codon_start=1
/product="atypical PKC isoform-specific interacting
protein long variant"
/protein_id="AAK27891.1"
/db_xref="GI:13491610"

/translation="MKTVCEGRTRVVPCGDGHNKVFSLIQAVTRYRKAIAKDPWY
WIOVRLHGHGGIILDDIIDLVDADKDLVAVEDDDPHGGDGTASSTGOSPE
IFGSELTNNVSAFOPIATSEIETPVLNAMPILHRSDDPALLIGLSTSVSDNF
SSEPSRKNPTRMSTAGTFLKONTAGSPKTCDDKDEYNSLPDITNSMNOFQDNA
RSLSLASHPMGKMLEKQEDDEGTEDNSHVEFVGHADGLEHLPNSLDDMYKVE
VPNDGPIGLHVPPSARGKTLGLVYRLKKGAKAHENLFRNDCIVRIDDLNLR
RFEQAOHMEFQAMRTPLIMHVVPANKEQYEDLSQSEKNYSSRSPSOYIDNR
SVNSAGLHTVQAPRLNHPPEIDSHSLPHASPSKPPSPAPASQONVSTVSSG
YNTKIGKRLNIOAKKREGLEFSTSDVIGGSADIVYKNILPRGAIIDGGLTKG
DLLEFVNGVDLVGSOEEVSLRSTMEGVSLVPRQDAHPRLNAPSPQMOLP
KETAEDEDIVLPDGTREPLFEVPLINDSSALGYSVGNRSKNEHADIGLIVKSI
INGGASKIDRLRVNDOLIAVNGSLKGTNMDMELTNRSMSTEGNKMIMOLIVAR
RISGCNELKSPSPPEPLPTELALDRERISLISLHGLEDESPRMAALSRING
KIQSLPVNMPQDVTIIEEDRLPVLPHSLSDSSSSHDVGTADATYMAKALS
DSACSSLPDVPVLAOFREGFROSSEKTKQSPASOLDPFTKTRKSMOLDIAD
ETKLNIVDDKAGSPSRDVGPSLKKSSLSLSLQTAFAVTTNGDIPRRPRRIIR
GRGNESEFRAIDKSDKPAVDDDEEMETLEEDTESSSSGSRSSVSTASDPSHSE
ROMNGONEKGRDKDKTKGKEKKKROKEMKAKKMLGLDMPHHRKRD
KIEKTGKIKIOESTSEERIIMKQEDERLOATREHREQAERYAEIODEHRTG
CDDPLMTGVSYSTEGSMALNARQSPREGHMDALYAQVYKPNRNSRSPVDSNRSTPS
MHDRIOIRKQEDAKODEVEDERRTYSEQPMVNPAPATQGRHSVSEVOMOR
QOEERSSQOARQVSLPQSRKNSASVSDSSOMVNPSPGQSARENDRSSVQGS
RNGVIGGHFNARVMELOELRROEORRKOOKKPPSPGSPNSYKRYQDPSTAP
PKGPRDQVPPSPSOVARLNRLOTPKGRFYS"

BASE COUNT 1623 a 1395 c 1581 g 1358 t 1 others

Query Match 85.2%; Score 4848; DB 9; Length 5958;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 5668; Conservative 0; Mismatches 2; Indels 10; Gaps 2;

QY 1 atgaagaatgaccgtgcttcgcgagcaaccgggtgctgctgcgtagcgaggagcaac 60
DB 288 ATGAAAGTGAACCGTGTCTTGCGACGACCGGGTGTCTGCGTGCAGGCGGCAC 347
QY 61 atgaagaatgaccgtgcttcgcgagcaaccgggtgctgctgcgtagcgaggagcaac 120
DB 348 ATGAAAGTGAACCGTGTCTTGCGACGACCGGGTGTCTGCGTGCAGGCGGCAC 407
QY 121 gatccaactactgagatagcagtgatcgcttggaacatgagatgaggaataactagac 180
DB 408 GATCCAAACTACTGAGATACAGGTGCAATCGCTTGGAACATGAGATGAGGAATACTAGAC 467
QY 181 cttagatgacatctcttgtagttagaagacagataaagaacagactgtagcaagtgttagat 240
DB 468 CTTAGATGACATCTCTTGTGTATGTAAGACAGATAAAGACAGACTGTGTACAGGTTTGAAT 527
QY 241 gagcagagatccacatcacaggaagtgatgcaaccagtgcaagttccacgggtgtaaccagac 300
DB 528 GAGCAGAGATCCACATCACGAGAGTGTATGACACAGTCCAGTCCAGGCTACCCAGAGAC 587
QY 301 ccagagagatatgtgtagtgaagcttgagcacaacaatgctcagccttcaagcttaacca 360

DB 588 CCAGAGATATTTGGTAGAGAGTGGACCAACAATGCTCAGCCTTACACCTTACCAC 647
QY 361 gaaacaagtgaattgaagtacacacttcacacttccttcgagcaaatatgctctcaatgt 420
DB 648 GCAACAAGTGAATTTGAGAGTACACACTTCACTCCAGCAAAATATGCTTCATGTT 707
QY 421 cgaagcagtagtgaacccagctctaattgctctccacttcctgtgcagtgtaagtaattt 480
DB 708 CGAGCAGTATGAGACCCAGCTCTAATTGGCTCTCCACTTGTGATGATGATTAATTT 767
QY 481 tctctgaagagccttcaaggaataatcccaacagctgtgtcaacaacagctgtctctc 540
DB 768 TCTCTGAAAGCCCTTCAAGGAAATCCACACAGCTGTGTAACAACAGCTGGCTTC 827
QY 541 aagcagaacactgctggaagtcctaaacactgagcagaggaagaagtgaaactacaga 600
DB 828 AAGCAGAACTGCTGGAGAGTCTTAAACCTCCACAGAGAAAGATGAAACTACAGA 887
QY 601 agcctcccgaggatcactagtaactggtcctaaccatltcaagagagaatgtctcgtc 660
DB 888 AGCTCCCGGGGATACTAGTAACCTGGTCAACCAATTTGAGAGACATGCTCGCTC 947
QY 661 tctctgagtgccagtcaccccaatgtgtggaagtgctgtggaagaacaagaacagatgag 720
DB 948 TCTCTGAGTCCAGTCAACCAATGAGTGGCAAGTGGCTGGAAGAAACAGATGAG 1007
QY 721 gatggagaagaagatgaacagtcgltgtgaacccgtgttgagatgctgacacaggttg 780
DB 1008 GATGGGACAAAGATATACAGTCTGTTGAACCTGTTGACATCTGACACGGGTTTG 1067
QY 781 gacatatacccaactttctctgtagatgatagtgaagctcgtagaagtcaccaagat 840
DB 1068 GAGCATATACCAACTTTCTTGATGATGATGATGATGATGATGATGATGATGAT 1127
QY 841 ggaagcctctggaatccatgtgagtccttcagtgctcgaagcgaggaacccctggag 900
DB 1128 GAGAGGCGCTGGGAATCATGTAGTCCCTTCAAGTCTCGAGCGCGGAAACCTCGGG 1187
QY 901 ttatttgtaaaagatattgagaagaagtgtaagctgaacatgaataatcttttcgtgag 960
DB 1188 TATTTGTAAGAGATGAGAGAAAGCTGTAAAGCTGAACATGAAATCTTTGCTGAG 1247
QY 961 aatgattgcatgctcaggttaatgtagtcgaccccttcgaatagaagatttgaacaaga 1020
DB 1248 AATGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1307
QY 1021 caacatagtcttcgcaagcagatgctgtaacacccatcatgttgcacatgtgctgca 1080
DB 1308 CAACATATGTTTCCGCAACCCATGCTGACCCATCATTTGTTCCATGTTCTCTGCA 1367
QY 1081 gcaaatagaagcagtagaacaactatcccaaatgtgagaagaacaataactatcaaga 1140
DB 1368 GCAATTAAGACAGTATGACAACTATCCAAATGTGAAGAAATTAATTAATCAAG 1427
QY 1141 ggttttagccttgacagccagatataatgacaacagagtgtagaagtgagaggttcaac 1200
DB 1428 CTTTTAGCCCTGACAGCAGATATATGACAACAGAGATGTGAACAGGCTTCAC 1487
QY 1201 acggtgagagagaccccgagatgaaaccaacggcctgagagatagactcactcaaga 1260
DB 1488 ACGGTGACAGAGACCCCGAGTGAACCCCGCTGAGAGATGAGACTCACTCAAGA 1547
QY 1261 ctactcctaaagcacaacccctcgaggaacaacacatccgttcagactcggaactcag 1320
DB 1548 CTACTCTAAGACGACACCCCTCGGGAACACACCATCCGCTTCCAGCTCGGCAAC 1607
QY 1321 aatgtatttagtgaagctgtaagcagtggttatataacccaataaataaggaagagct 1380
DB 1608 AATGTATTAGTGAAGCTGTAAAGCAGTGTATTAACCAAAAATAAGCAAGAGCTT 1667
QY 1381 aatatacagcttaagaaggtacagaaggttttgatcagcatcactccagaagatgta 1440

Db 1668 AATATCCAGCTTAAAGAAAGTACAGACGGTTGGATTGAGCATCTTCCAGAGATGTA 1727
Qy 1441 acaataagg tggctcagctcccaaltcctatgtaaaaacatctcccccggggcgccatt 1500
Db 1728 ACAATAGGTGGGTGAGCTGCATCTATGTGAAAAACATTCTCCCGGGGGGGGCCANT 1787
Qy 1501 caggatggcgaacttaaggcaggagacgaacttataagtaataatggatagattagtg 1560
Db 1788 CAGGATGGCCCACTTAAGCGAGAGACAGACTTATAGAGTAAATGGATAGATTAGTG 1847
Qy 1561 ggcgaatcccaaggaggagattggttcctcgctgttggagaagacccaagaatggagaactgtg 1620
Db 1848 GGCAAATCCCAAGAGGAAGTGTCTTGGCTGTGAGAACCCCAAAATGGAAGAACTGTG 1907
Qy 1621 agccttcgtcttcgcgaagaaagacgctccacccaaggaaactgaatcacagcca 1680
Db 1908 AGCCTTCTGGTCTTTCGGCAGGAACAGCCTTCCACCCAGGGAATGATACAGACCA 1967
Qy 1681 agccaagatgcagatcccaaaagaaacgaagcagaagaatgaggataatgttcttaacct 1740
Db 1968 AGCCAGATGCAGATTCCAAAAGAAACGAAGCAGAAATGAGATATTGTCTTACACT 2027
Qy 1741 gatggcacaagggaattctgacatttgaatgccacttaataatgacagatctcagagc 1800
Db 2028 GATGCAACAGGGAAATTTCTGACATTTGAAGTCCCATTAATGATTCAGGATCTCAGGC 2087
Qy 1801 ctgtgtctcagtgctcaaaagtaaccggtccaaagaaacacagcagatttggaaactctt 1860
Db 2088 CTTGGTGTCACTGTCAAAAGTAAACCGGTCAAAAGAAACACGCGAGATTGGGAATCTT 2147
Qy 1861 gtcaagtcattatataatgtagggagcagcatctaaagaatgaaaggctcggtgtagatg 1920
Db 2148 GTCAAGTCCATTATTAAGAGGAGAGACACATCTAAAGATGAAAGGCTCGGGTGAATGAT 2207
Qy 1921 caactgatagcaagtaaatggagaatccctgttggcgaagaaacaaacaagatgccatgaa 1980
Db 2208 CAACTGATAGCAGTAAATGAGAAATCCCTGTGGCAAGCAACCAAGATCCCAATGGGA 2267
Qy 1981 accctaagaaggtctatgtctacttgaaggcaataaagcagaaatgcatcaacttattgt 2040
Db 2268 ACCCTAAGAAGGTCTATGTCTACTAAGCAATTAACGAGAAATGATCCAGTTATTGT 2327
Qy 2041 gcaaggaagaataagcaagtgcgaatgaatgaatcaacttggagacccctggagcctgag 2100
Db 2328 GCAAGGAATAATGACAACTGCAATGACGTGAAGTCACTGGGAGCCCTTGACACTGAG 2387
Qy 2101 ctgcccatttgaagaacggttggatagatagaagaagaagaatttccattccctctacagt 2160
Db 2388 CTGCCCATTTGAACACGCGTTGGATGATAGAGAACGAAAGAAATTTCCCATTTCCCTTACAGT 2447
Qy 2161 ggaattgaggggcttgaataatcgcccagcagaataatgctgcctcctagtagaataatggt 2220
Db 2448 GGGATTTGAGGGGCTTGATGAATCGCCCAAGCAAAATGCTGCCCTCAGTAAGTAATG--- 2504
Qy 2221 gagtcaaggttaataacacagctgtccctctacagtgaataatgccccaaagatgaacgttcatt 2280
Db 2505 -----GGTAAATACCACTCTCCCTACAGTAATATGCCCAAGATGACACTGTCAATT 2558
Qy 2281 atagaagaatgcaggttgcaggtgtctctccacaatctcttgcagcaatccctctccagc 2340
Db 2559 ATAGAAAGATGACAGGTTCCAGTGTCTTCCACATCTCTGACACAGTCCCTTTCACAGC 2618
Qy 2341 tcccatgtatgtatgtaggttttgtgaacgagatgtgttacttggccaaggtcgaatc 2400
Db 2619 TCCCATGTATGATGTGGGTTTGTGTGACGGCAGATGCTGTACTTGGCCAAAGCTGCANTC 2678
Qy 2401 agtgaattcagcagactgtctcttgaatccagaatgttgaatccagttctgtcttccaaga 2460
Db 2679 AGTGAATTAGCCGACGTGCTTTGTGATCCAGATGTTGATCCAGTTCTGTCTTTCAAGCA 2738
Qy 2461 gaaggaatttgcagctcagaagatattcagaaaaaacgacaaaagcaatttccaatgacagt 2520
Db 2739 GAAGGATTTGAGAGCTCAGAGTATGTCAAGAAACGCAAAAGCAATTTTCAGATCCAGT 2798

Qy 2521 caatttgaatttgcgttaaaaaacacgaaataatcaaaaagcatgtatttaagtaatactgacag 2580
Db 2799 CAATTGGATTTGTTAAACACAGAAATCAAAAAGCATGATGATTTAGTATAGTGCAGAG 2858
Qy 2581 actaaactcaatacagtgtagatgaacgaagaagcaggtctctccaagcagagatgttgctct 2640
Db 2859 ACTAAATCTCAATACAGTGTGATGACCGAAGAACAGGTTCTCCACGACGAGATGTGGTCTCT 2918
Qy 2641 tcccctgggtcttaaaagaatgaagctcggttggagagcttgcacaacgcaatttgcagagtg 2700
Db 2919 TCCCTGGGCTTAAGAAAGTCAAGCTTATGAGAGTCTCCACACCCGATTCGCCGAGTG 2978
Qy 2701 actttgaatggagataatctcttccatcgctccacggcccgagataatcagagcaagggga 2760
Db 2979 ACTTTGAATGGGGATATTTCTTTCATCGCTCCACGGCCCGGATATATCAGAGCAGGGGA 3038
Qy 2761 tgcgaatgagcttcaagcttgcacatgcgaacaaacttataatgaataaccgctgtagatg 2820
Db 3039 TCCCAATGACAGCTTCAGACCTGCCATCGCAAAATCTTATGATTAACCCCGGTAGATGAT 3098
Qy 2821 gatgtatgaaggatggagaccttggagaagaacagaagaagaatccaagatccaaggaga 2880
Db 3099 GATGATGAAGGCAATGGAGACCTTGGAGAAAGACAGAAAGAAATTCAGATCAGGAGAA 3158
Qy 2881 gagtctgtatccacagccagtgatcagccttccactctcttggagagacaaatgaaatgga 2940
Db 3159 GAGTGTGATCCACAGCCAGATGATCAGCTTCCCATCTCTGAGAGACAAATGATGGA 3218
Qy 2941 aaccaagaagaagtgtagaagactgtagaanaaagaataaaactggaanaagaagaag 3000
Db 3219 AACCAAGAAAGGTGATGATGAGACTGTAGAAAAAAGATTAACCTGGAAGCAAAAGAG 3278
Qy 3001 aagatagaagataaaggagaagataaataatgaagaacgaagaaggaaatgcttgaaggcttg 3060
Db 3279 AAGATGAGATGAAGAGAGAAAGTAATAATGAAAGCCAAAGGAAGGAATGCTGAAGGGCTTG 3338
Qy 3061 ggaagacatgttcaggttcttgcaaaacatcgaaaagatgaaagaatltgaanaacggttaaa 3120
Db 3339 GGAGACATGTTCAAGTTTGCAGAAACATCGAAAAAGATGACAAAGATTGAAAAACGGGTAAA 3398
Qy 3121 ataaaaatacaggaatccttatacataagaagaaggagaagatacgaatgaacgaagacag 3180
Db 3399 ATAAAAAATACAGGAATCCCTTACATCAGAAAGAGAGAGATACGAATGAACGAGGACAG 3458
Qy 3181 ggaagatccaagccaaactcgaagaatttgaagaacggaacagctggaagcgttgaact 3240
Db 3459 GAGAGATTTCAAGCCAAACCTCGAGAAATTTAGGAAACGACAGCTCGAGACCTGACTAT 3518
Qy 3241 gctgaaatccaagatttcaatcggaacttggctgtgaatgaatgaaatgtataatggagga 3300
Db 3519 GCTGAATAATCAAGATTTTTCATCGAGACATTTGGCTGTGATGATGATTAATGATGGGGGA 3578
Qy 3301 gttcttcttatagaaggttccatgtctcgaagctagaactcagaagcccaagaagaagg 3360
Db 3579 GTTTCTTCTTAAGAAAGTTCCATGTGCTCAACGCTAGACCTCAGAGCCACGAGAAAGG 3638
Qy 3361 catatgatgatgtcttgtatgcccgaagtcagaagaacgaggaatttccaaacctcaact 3420
Db 3639 CATATGATGATGATGCTTTGTATGCCCAAGTCAAAGAAACCCGGGAATTCAAACCTTACCT 3698
Qy 3421 gtagacagtaacagatccaactcctagcaatcatalgtatcagatatacagcttctgagcaaga 3480
Db 3699 GTAGACAGTAACAGATCAAACTCTAGCATCATGATCGGATACAGGCTGAGGCAAGAA 3758
Qy 3481 ttcaagcaagaacgaagaatgaagaatgttagaagaatcgttcgagcagactatgcttggag 3540
Db 3759 TTTTCAGCAAGCAAAACCAAGATGAAGATGTAGAAAGATCGTCGGCGACCTATAGTTTGGAG 3818
Qy 3541 caaccttggccgaacgacagcgcgagcagagagagggcgagcactcgggtgtccgttggag 3600
Db 3819 CAACCTTGCCCAAGCAGCAGCGCCGGCGACGACAGAGCGGGCGACACTTCGGTGTCCGTGGAG 3878

[illegible][illegible]

ACCESSION AF467002
VERSION AF467002.1 GI:18568343
KEYWORDS
SOURCE
ORGANISM human.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 4071)
TITLE Gao, L., Macara, I.G. and Jobery, G.
JOURNAL Multiple splice variants of PAR3 and of a novel related gene, PAR3L, produce functionally different proteins
REFERENCE 2 (bases 1 to 4071)
AUTHORS Gao, L., Macara, I.G. and Jobery, G.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of Virginia, PO Box 800577, Charlottesville, VA 22906, USA
FEATURES
SOURCE location/Qualifiers
1..4071 /organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1..4071
/gene="PAR3"
/note="PAR3"
1..4071
/gene="PAR3"
/note="PAR3"
/note="PAR3: multi-PDZ domain containing protein; alternatively spliced"
/codon_start=1
/product="partitioning-defective 3 protein splice variant a"
/protein_id="AAL76042.1"
/db_xref="GI:18568344"
translation="MKVTVCGRTRRVVPCGDGMKVFSLIQAVTRKAIADPNY
WIOYHREHGDGILDLIDLCDVADDKRIVAFEDODPHHGGDGRSASSTGOSPE
IPSGELCTNNVSAROPYOATSEIPEVPSVLARNPLRYRSDPALLIGLSISYSDNS
SSEPSRKNTPTRMSTAGFLKONTAGSPKCTDRKDEKYNRSLPHDTSWAKQFORDNA
RSSLSASHPWGWKLEKQEDDEDEEDNSVEVEGHADGLEHLPNSLDDMKLVE
VPNDGSLGHVVEFSARGRTGLLVKRLKGGKAHEMLFRNDCIVRINDDLRN
RREOAOHMEFOARARPTIIFWHPVPAANKEDYEQLSQSEKNYSRSPSQYIDNR
SVNSAGLHTVORAPRLNHPQIDSHSRPLSAHPSGKPPAPASADONVSTVSSG
YNTKKIKRLINIOIKKCTEGIGFSTSRDVTIGSAPVYVNTLPRCAIODGLKAC
DLLEVNGVULVKSGSEVYSLRSTYMEGVSLVYRQEDAFPRELNAEPMQMDP
KETAEDEDIVLPDGTREFLPTEFVPLNDSGSLGVSVKNRSEKEMHADGIVKSI
INGGAASKDRLARVNDQILVANGSLGKIQMDAMETLRSMSTEGNKGMIQIVAR
RISKCNELKSPSGPELPIETALDRERISHSISIGLDESPSRNALSRIMG
ESGYOLSPYNNMPDQDVIIIEDRLPVLPPLPHLSDOSSSSHDVGYTTAGAWAKA
AISDASCSLSPVDVPIARQRRGROSMSEKRTKOPSDASQLDPVYTKRSKMDLG
IADERTLNTYDDOKAGSPSRDVGPSLGLKSSSLSLQTAAYAEVTLGNDIFPHPRK
IIRGCGNESFRALIDKSYDKPAVDDEDEGETLEETESSRGRSRSVSTASQPSH
SLERQNGNOKGKIDRKDKTKGKERKDKDEKDKAKKMKLGLGDMFRGKIH
KDKIERTGKIKIESEFTSEERIRMKOEORIOAKREFREPAREDEYIODEFIR
TFGCDDELATGGVSYEGSMALNARPOSPRGHMDLVIAVKKPRNKKSPVDSNRS
TPSNHDIORLROFOAKODEVEDERARYSEOPPNAPAPAQOSRHSVYQMO
KORERESSQAKQROTSLPROSRRKNAVSQSDMONTSPGSGFQSAKKNPYTSI
QGSNGVLGGHGFARVMLETQELRLQERKREQOMKPPSESGPSNDSYKKVQDPS
YAPKGPFRQDVPPSPQVARNRLQTEPKGRPYs"
814..1074
/gene="PAR3"
/note="Region: PDZ domain"
1378..1638
/gene="PAR3"
/note="Region: PDZ domain"
misc_feature
/gene="PAR3"
/note="Region: PDZ domain"
1765..2028
/gene="PAR3"
/note="Region: PDZ domain"
BASE COUNT 1220 a 923 c 1079 g 849 t
ORIGIN

Query Match 70.7%; Score 4020; DB 9; Length 4071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgaaagtacccgtgtcttcggaacggaacccgggtgttcgtccgtgcggggagcgccac 60
DB 1 ATGAAAGTACCCGTGTGCTTCGAGACGACCCGGGTGTGTCGCGCGGGACGGCCAC 60
QY 61 atgaaagtlttaagcctcatccagcagcggtgtgacccgttacccgaaggccatcgccaag 120
DB 61 ATGAAAGTTLTAAAGCCTCATCCAGCAGCGGGTGTGACCCGCTACCGGAAGCCATCGCAAG 120
QY 121 gatccaactactggaatacaggtgtcgccttggaacaatggaatggaataactatagc 180
DB 121 GATCCAACACTACTGGATACAGGTGCATCGCTTGGAACATGGAGATGAGCAATACTAGAC 180
QY 181 ctgtatgacattctcttgtatgtatgacagaatgaaagacagactggttagcagttgtat 240
DB 181 CTGTATGACATTCTCTTGTATGTATGACAGACAGATAAAGACAGACTGTTAGCAGTGTGAT 240
QY 241 ggcgcagatccacatcagcaggaagtgtatgtgcacacagtcgcaagttccaggtaccagagc 300
DB 241 GGCAGGATCCACATCACGAGGTGTATGTGCACACAGTCCAGTCCACGGGTACCCAGAC 300
QY 301 ccagagataattgttagttagtgccttgacacaaatgctcagcctttagccttaccaa 360
DB 301 CCAGAGATATTGTGTAGTGTAGTGCCTTGACACCAATGTCTCAGCTTTAGCCTTACCAA 360
QY 361 gcaacaagtgaatltgaagltcacacactcagtccttcgagcaataatgctcttcagt 420
DB 361 GCAACAAGTGAATLTGAAGLTCAACACTCAGTCCTTCGAGCAAAATATGCTCTTATGTT 420
QY 421 cgaagcagtagtgcacccagccttaattgtgctctccaattctgtcagtgatgtaattt 480
DB 421 CGAAGCAGTAGTGCACCCAGCCTTAATTGTGCTCTCCACTTGTCTCAGTATGTAATT 480
QY 481 tctctgaaagagccttcaaggaataatccacacagctgtgtcaacaacagctgctcc 540
DB 481 TCTCTGAAGAGCCTTCAAGGAATAATCCACACAGCTGTGTCAACAACAGCTGCTTCTC 540
QY 541 aagcgaacactgtctggaagtccttaaaacttcgaacgaagaagaatgaaactacaga 600
DB 541 AAGCGAACACTGTCTGGAAGTCTCTAAACTCTGAGACGAGAAAGATGAATAACTACAGA 600
QY 541 aagcgaacactgtctggaagtccttaaaacttcgaacgaagaagaatgaaactacaga 600
DB 541 AAGCGAACACTGTCTGGAAGTCTCTAAACTCTGAGACGAGAAAGATGAATAACTACAGA 600
QY 601 agcctcccgcgagatactatgtaactgtgttaacaaattccagaagagaactgtcgctg 660
DB 601 AGCCTCCCGGAGATACATACTGTACTGTCTTAACAAATTTCAAGAGACATGCTCCTCG 660
QY 601 tctctgaagtcgaagtcacccaatgtgtgcaagtcgtgagaataaagaacagagatgag 720
DB 601 TCTCTGAGTCCAGTCAACCAATGTGTGCAAGTGTGAGAAACAAGACAGATGAG 720
QY 721 gatggagacagaagagataacagtcgtgttgaactgttgcacatgtctgcacacggattg 780
DB 721 GATGGAGACAGAAGAGATTAACAGTGTGTTGAACGTGTTGACATGCTGCACACGGATTG 780
QY 781 gagcatatacccaacttctctgtatgataatgttaagtcgttagaagtccecaagat 840
DB 781 GAGCATATACCCAACTTCTCTGTATGATATGTAAACCTGTATGAATCCCAACGAT 840
QY 841 ggaagagcctcgtggaatccatgtagtgccttcagtcgcgcgaagcggagaaacccctgggg 900
DB 841 GGAAGGCCCTCTGGGAATCCATGTAGTGCCTTTCAAGTCTGAGGCGGAGAACCCCTGGGG 900
QY 901 ttattagtaaaacgaatttggaagaaggtgtgtaaagcttgaacatgaaatcttttcgtgag 960
DB 901 TTATTAGTAAACGATTTGGAAGAAGTGTGTAAGCTGAACATGAAATCTTTTTCGTGAG 960
QY 961 aatgattgcatgtcgaagataatgataatgacgaacttcgaaataagaagaatttaacaagca 1020
DB 961 AATGATTGCAATGTGAGGATTAATGATGCGACCTTCGAAATAGAAAGATTGAACAAAGCA 1020
QY 1021 caacatatttgcgcaaacgcatgctacaccatcatttggttccatgtgttctctgca 1080
DB 1021 CAACATATTGTCGCCAACGCATGCTACACCATCATTTGGTTCCATGTGTGTTCTCTGCA 1080

OY 1081 gcaataaagagcagatgtaacacactaaccacaaagtgaagaagacattactatccaaagc 1140
DB 1081 gcaataaagagcagatgtaacacactaaccacaaagtgaagaagacattactatccaaagc 1140
OY 1141 cgttttagccctgacagccagctatattgcaacaagaagtgtgaacagtgcaaggtctaac 1200
DB 1141 cgttttagccctgacagccagctatattgcaacaagaagtgtgaacagtgcaaggtctaac 1200
OY 1201 accgttgagagagcaaccccgactgaacacaccccgctgagacagatagatcttcacaa 1260
DB 1201 accgttgagagagcaaccccgactgaacacaccccgctgagacagatagatcttcacaa 1260
OY 1261 ctacccatagcgacacacccctcggaanaacacacatccgctccagcctcgcaactcag 1320
DB 1261 ctacccatagcgacacacccctcggaanaacacacatccgctccagcctcgcaactcag 1320
OY 1321 aatgtatagtaagcagctgtaagcagtggtatatacaccaaaaaataagcaaggcct 1380
DB 1321 aatgtatagtaagcagctgtaagcagtggtatatacaccaaaaaataagcaaggcct 1380
OY 1381 aatatacagcttaagaagaagctacagaaggtttggatctcagcatcctccagatgta 1440
DB 1381 aatatacagcttaagaagaagctacagaaggtttggatctcagcatcctccagatgta 1440
OY 1441 acaataagtgagctcagctcacaatctatgtgaanaacattcccccggggggcgagcatt 1500
DB 1441 acaataagtgagctcagctcacaatctatgtgaanaacattcccccggggggcgagcatt 1500
OY 1501 caggatggcagacttaaggcagcgagacagactatagaggtaaatggaattagtg 1560
DB 1501 caggatggcagacttaaggcagcgagacagactatagaggtaaatggaattagtg 1560
OY 1561 ggcacaaatcccaagaagaagtgtctgcgtgtgaagaagcacaagaatggaagaactgtg 1620
DB 1561 ggcacaaatcccaagaagaagtgtctgcgtgtgaagaagcacaagaatggaagaactgtg 1620
OY 1621 agccttcgtctcttcgcccaggaagagcctccaccacaaaggaaactgaatcagagcca 1680
DB 1621 agccttcgtctcttcgcccaggaagagcctccaccacaaaggaaactgaatcagagcca 1680
OY 1681 agccagatgcagattcccaaaagaagaagaagcagaagatgaagatattgtcttaacct 1740
DB 1681 agccagatgcagattcccaaaagaagaagaagcagaagatgaagatattgtcttaacct 1740
OY 1741 gatggcaccaggaattcttcgacatttgaagtcccaacttaagtltcaagatcgcagagc 1800
DB 1741 gatggcaccaggaattcttcgacatttgaagtcccaacttaagtltcaagatcgcagagc 1800
OY 1801 cttgtgtcagtgtcaaaagttaaccggtcaaaagaagaacccagagatttgggaactctt 1860
DB 1801 cttgtgtcagtgtcaaaagttaaccggtcaaaagaagaacccagagatttgggaactctt 1860
OY 1861 gtcaagtcacatattaatgagagagcagcatctaaagaatggaagcttcggtgtaattgt 1920
DB 1861 gtcaagtcacatattaatgagagagcagcatctaaagaatggaagcttcggtgtaattgt 1920
OY 1921 caactgataagcagtaaaatggaatccctgtgtggaagaacaaacaaatgcatgtaa 1980
DB 1921 caactgataagcagtaaaatggaatccctgtgtggaagaacaaacaaatgcatgtaa 1980
OY 1981 accctaagaaggtctactgttactaaggaacataaagaagaatgtaaccagcttaattgt 2040
DB 1981 accctaagaaggtctactgttactaaggaacataaagaagaatgtaaccagcttaattgt 2040
OY 2041 gcaaggaagaataagcaagtgtcaatgagctgaagtcacctcggaagcccccgtgagctag 2100
DB 2041 gcaaggaagaataagcaagtgtcaatgagctgaagtcacctcggaagcccccgtgagctag 2100
OY 2101 ctgcccattgaacacagcgtttgatgatagagaagaagaatctccattccctctacagt 2160
DB 2101 ctgcccattgaacacagcgtttgatgatagagaagaagaagaatctccattccctctacagt 2160
OY 2161 gggatctgaggggcttgatgaatcgcacaggaatgtgtccctcagtaggataatgggt 2220
DB 2161 gggatctgaggggcttgatgaatcgcacaggaatgtgtccctcagtaggataatgggt 2220

DB 2161 gggatctgaggggcttgatgaatcgcacaggaatgtgtccctcagtaggataatgggt 2220
OY 2221 gaatcagtaaatatacagctgtccctcagtaggataatgccccagaatgacatgtaalt 2280
DB 2221 gaatcagtaaatatacagctgtccctcagtaggataatgccccagaatgacatgtaalt 2280
OY 2281 atagaagaatgaagaaggttcagctgctctcccaatcctctctgaacagctcctccagc 2340
DB 2281 atagaagaatgaagaaggttcagctgctctcccaatcctctctgaacagctcctccagc 2340
OY 2341 tcccatgatgatgtgggtttgtgtacggcagatgtcgtacttggtggcgaagctgtaac 2400
DB 2341 tcccatgatgatgtgggtttgtgtacggcagatgtcgtacttggtggcgaagctgtaac 2400
OY 2401 agtatttaagccagctgctcttgagtcagatgtgatcagtttcttcttcaaga 2460
DB 2401 agtatttaagccagctgctcttgagtcagatgtgatcagtttcttcttcaaga 2460
OY 2461 gaagatcttgagctgaagatgtaagtaagaagaacgcaagaacatcttcagatgcagt 2520
DB 2461 gaagatcttgagctgaagatgtaagtaagaagaacgcaagaacatcttcagatgcagt 2520
OY 2521 caattgagatctcgttaaaacagcaaaatcaaaagaatgattggtatagctgagag 2580
DB 2521 caattgagatctcgttaaaacagcaaaatcaaaagaatgattggtatagctgagag 2580
OY 2581 acttaaatcaatacagtgatgatacagaagaagaagttctccagcagaagatggtgctc 2640
DB 2581 acttaaatcaatacagtgatgatacagaagaagaagttctccagcagaagatggtgctc 2640
OY 2641 tccctggtctgaagaagtcagctcgtgtgagagtcctgcaagacagcaggttgcagagtg 2700
DB 2641 tccctggtctgaagaagtcagctcgtgtgagagtcctgcaagacagcaggttgcagagtg 2700
OY 2701 actttgaatgggagatactcttcctcatgctcaagccggcgagataatcagaaggagga 2760
DB 2701 actttgaatgggagatactcttcctcatgctcaagccggcgagataatcagaaggagga 2760
OY 2761 tgcattgagaagcttcaagactgacatcgacaacatcttatgataaccggcgtagatgat 2820
DB 2761 tgcattgagaagcttcaagactgacatcgacaacatcttatgataaccggcgtagatgat 2820
OY 2821 gatgtgaagcagatggaacacttggaaagaagacacagaaagaatcgaatcagaaggaga 2880
DB 2821 gatgtgaagcagatggaacacttggaaagaagacacagaaagaatcgaatcagaaggaga 2880
OY 2881 gattctgtatccacagccagtgatcagccttcccaactctcttgagaagaataatgtaga 2940
DB 2881 gattctgtatccacagccagtgatcagccttcccaactctcttgagaagaataatgtaga 2940
OY 2941 aaccaagaagaagtgataagactgataagaaagaalaagaactggaagaagaagaag 3000
DB 2941 aaccaagaagaagtgataagactgataagaaagaalaagaactggaagaagaagaag 3000
OY 3001 aaagatgagaataagaagaagaataaataagaaagcgaagaagaagatctgaagggtctg 3060
DB 3001 aaagatgagaataagaagaagaataaataagaaagcgaagaagaagatctgaagggtctg 3060
OY 3061 ggaagacatgtcaggttttgcaaacatcgaaagaatgacaagaatgagaanaacggtgtaa 3120
DB 3061 ggaagacatgtcaggttttgcaaacatcgaaagaatgacaagaatgagaanaacggtgtaa 3120
OY 3121 atataaatacagaagaatccttatacagaagaagaagaatgacaagaatgagaanaacggtgtaa 3180
DB 3121 atataaatacagaagaatccttatacagaagaagaagaatgacaagaatgagaanaacggtgtaa 3180
OY 3181 gagagatctcaagcacaacatcgagaatggaagaagaagaatcgaagcgtgagactat 3240
DB 3181 gagagatctcaagcacaacatcgagaatggaagaagaagaatcgaagcgtgagactat 3240
OY 3241 gctgaatcagaatcttcacatcgacatttgctgtatgatgattatgtagtgggga 3300
DB 3241 gctgaatcagaatcttcacatcgacatttgctgtatgatgattatgtagtgggga 3300

[illegible]

Dh 2332 TCCCATGATGATGTTGGGTTTGTGACGGCAGATGCTGTACTTGGGCCAAGGCTGCAAATC 2391
Qy 2401 agtgaatcaagcagcagctctcttgagtcagagatgtgaatccagttcttgcttcttcaagca 2460
Dh 2392 AGTGATTCAGCCGACGCTGCTTTGAGTCCAGATGTGATCCAGTTCCTGCTTTTCAAGCA 2451
Qy 2461 gaaggaatttgacgtcagagatgtcagaanaaagcacaaagcaatcttcagatgacagt 2520
Dh 2452 GAAGGATTTGGACGTCAGAGTATGTCAAAAAACGCAAAAGCAATTTTCAGATGCCAGT 2511
Qy 2521 caattggatttcgttaaaacagaaaatcaaaagcatgagattagatgatgtgagagag 2580
Dh 2512 CAATTTGATTTGCTTTAAACACGAAATCAAAACCATGATGATTAAGTATGCTGACGAG 2571
Qy 2581 actaaaccaaatacagtgatgacagcaagaagaagttctccacagagatgtggtctct 2640
Dh 2572 ACTAAACCATATACAGTGGATGACGAGAACGAGGTTCTCCAGCAGAGATGTGGGTCTCT 2631
Qy 2641 tccctgggtctgaagaagtcgaagctcgttggagagcttcgagaccgagcttgccagagt 2700
Dh 2632 TCCCTGGGTCTGAAGAAATCAAGCTCGTTGGAGAGTCTGCAGACCGCAGGTTGCCAGGTG 2691
Qy 2701 actttgaaatgggagatattcccttccatcgttcaagcgccgagataatcaagggcagggga 2760
Dh 2692 ACTTGAATGGGGATATTCCTTTCCATCGTCCACGCGCGGATATATCAGAGCGGGGA 2751
Qy 2761 tgcagtgagagcttcaagagctgcacacaaatcttaagtaaaacccgagtaagatgat 2820
Dh 2752 TGCATATGAGACTTCAGAGCTGCCATGCACAATCTTATGATTAACCCGCGGTAGATGAT 2811
Qy 2821 gatgaatgaagcattgagagaccttggagaagaacacagaagaagtccaagtccaaggaga 2880
Dh 2812 GATGATGAAGGACATGAGAACCTTGGAAAGAACACAGAAAGAAAGTTCAAGATCAGAGGAGA 2871
Qy 2881 gagctgtgataccaagccaagtgatcagcttccacactctcggagaagcaaatgatatga 2940
Dh 2872 GAGTGTGATCCACACCGAGATGATCGCTTCCACTCTCTGAGACGACAAATGATGGA 2931
Qy 2941 aaccaagaagaagtgataaagacttgatagaaaagaataaaactcggaagaagaagaag 3000
Dh 2932 AACCAAGAAAGGTGATTAACATGATGAAAAAAGATTAACACTGGAAGAAAAAGAAAG 2991
Qy 3001 aaagatagagataaggaagaagataaaatgaagaaccaagaagggaatgctgaaggcttg 3060
Dh 2992 AAAATATAGATTAAGGAAAGATTAATGAAGCCAAAGAGGGAATGCTGAAAGGCTTGG 3051
Qy 3061 ggaagacatgctcaaggttggcaaaacatcgaaaagaatgacaagaattgaaaaaaagtgtaa 3120
Dh 3052 GGAGCATGTTTCAGGTTTGGCAACATCGAAAGATGACAAAGATTGACAAAAACGGGTAAA 3111
Qy 3121 aataaaatacaggaatcccttatacatcagaagaagagagatacgaatgaagcagaagcag 3180
Dh 3112 ATTAATAATACAGGAATCCTTTATCATCAGAAAGAGAGAGATGAGATGACAGAGACAG 3171
Qy 3181 gagaggaatcaagccaaactcagagaatttagggagacgaagaagctcgagagcgtgactat 3240
Dh 3172 GAGAGGATTTCAAGCCAAACATCGAGAAATTTAGGGAAACAAAGCTCGAGAGCTGACTAT 3231
Qy 3241 gctaaatcaagaatttcacatcgagacatttgctgtgtgattgattatgattatggggga 3300
Dh 3232 GCTGAATTTCAAGATTTTTCATCGGACATTTGGCTGTGATGATGATTAATGATGGGGA 3291
Qy 3301 gttcttcttataagaagttcactgagctcacaagctataagctcaagaagcccaagaaagg 3360
Dh 3292 GTTCTCTTATGAAGGTTCATGGCTCTCAACGCTAGACCTCAAGGCCACAGAAAGGG 3351
Qy 3361 catatgagatgcttctgtatgcccgaagtcagaagaagccgaggaattccaacccctcact 3420
Dh 3352 CATATGATGATGCTTTTATAGCCCAAGTCAAGAGCCGCGGAATTCAAAACCTTCACCT 3411
Qy 3421 gtagacagtaacagatacaactccctagcaatcatgatacgatacaagcgtctgaggaagaa 3480
Dh 3412 GTAGACATTAACAGATCAACTCTTAGCAATCATGATCGGATACAGCGTCTGAGGCAAGAA 3471

Qy 3481 ttacagcaagaagaagaatgaagaatgtagaagaatcgctcgaggaactatagtttag 3540
Dh 3472 TTTTCAGCAAGCAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3531
Qy 3541 caaaccttgcggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3600
Dh 3532 CAACCTTGGCCCAACGACACGGCCGCGGACGACGACGCGGCGGACGCTCGGTGCTGGAG 3591
Qy 3601 gtcagatgcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3660
Dh 3592 GTGCAATGCAAGCGGACGCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3651
Qy 3661 tacagctctctgctcctcggaagaagaagaatgcaagctcgtctccagagactcttg 3720
Dh 3652 TACAGCTCTCTCTCTCGGAAAGAGAAATGCAAGCTCGGTCTCCAGAGACTCTTG 3711
Qy 3721 gagcagaactactcccttggaggagcttcaagatgccaagaagaaccagactctcc 3780
Dh 3712 GAGCAGAACTACTCCCTCGGAGAGGCTTCCAGAGTGCCTAAAGAGAACCCAGGTTATCC 3771
Qy 3781 agctlaacaaagctcaggaagaagctactcgtggagagacatgcttcaagcagagctcatg 3840
Dh 3772 ACCTACCAAGGCTCCAGAGAGGCTTACTGGAGAGACATGCTTCAACCCAGGTCATG 3831
Qy 3841 ctggaactcaggaagctccttcgccaagagacagagcgggaagaagcagacagatgaagaag 3900
Dh 3832 CTGGAAGCTCAGAGACTCTCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGATGAAGAAG 3891
Qy 3901 cagctccttcgaggagggccagcaactatgactcgtataagaagaatccagagaccagct 3960
Dh 3892 CAGCTCTCTCTCGAGAGGCGCCAGCAACTATGACTGTTATTAAGAAAGTCCAGAGCCCACT 3951
Qy 3961 tacgccccctcagaaggagccttcggaagaatgagccccctccctctgaagtttg 4020
Dh 3952 TAGGCCCTCCCAAGAGGCGCTTCCGAGAGATGTCCTCCCTCTCTCAGGTTGG 4011
Qy 4021 aggcctgaacagactcagactccttgagaagaaggagccctctatctctga 4071
Dh 4012 AGGCTGAACAGACTCAGACTCTGAGAAAGGAGGAGCCCTTCTATTTCTGCA 4062

RESULT 4
AF467004 3960 bp mRNA linear PRI 07-PEB-2002
LOCUS Homo sapiens partitioning-defective 3 protein splice variant d
DEFINITION (PAR3) mRNA, complete cds, alternatively spliced.
ACCESSION AF467004
VERSION AF467004.1 GI:18568347
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3960)
AUTHORS Gao,L., Macara,I.G. and Joberty,G.
TITLE Multiple splice variants of PAR3 and of a novel related gene,
PAR3L, produce functionally different proteins
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3960)
AUTHORS Gao,L., Macara,I.G. and Joberty,G.
JOURNAL Direct Submission
TITLE Submitted (10-JAN-2002) Center for Cell Signaling, University of
Virginia, PO Box 800577, Charlottesville, VA 22908, USA
FEATURES
source
1..3960
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3960
/gene="PAR3"
/note="PAR3"
1..3960
/gene="PAR3"

Oy 1501 caggatggcgaacttaagggcaggagacagacttaagagtaaatgaggttagatltag 1560
 Db 1501 CAGGATGGCCCACTTAAGGCGAGAGACAGACTTATAGAGGTAAATGCGATGATTTAGTG 1560
 Oy 1561 ggcgaatcccaaggaggaagtgttgcgtgtttgagaacacccaagaatggaaggaactgtg 1620
 Db 1561 GGCCAATCCCAAGAGGAAGTGTGCTGTGAGAACACCAATATGAAAGAACTGTG 1620
 Oy 1621 agccttcggtccttcgcgcagaagaagccttcacccaagggaactgaatgcagagcca 1680
 Db 1621 AGCCTTGGGTCTTTCGCCAGAGAAAGCCCTTCCACCAGGAAGCTGATATCAAGCCA 1680
 Oy 1661 agccagatgcagatcccaaaagaaagcagaagaatgaagatatgttcttaacct 1740
 Db 1661 AGCCAGATGCAGATTCCAAAGAAACGAAGCAGAAATGAGATATGTTCTTACACT 1740
 Oy 1741 gatggcaccaggaatctctgaacttgaagtcaccaatgaatcaggatcagagc 1800
 Db 1741 GATGGCACCGGAATTTCTGCATTTGAGTCCCATTAATGATTCAGGATCTCAGGC 1800
 Oy 1801 ctgtgtcagtgtaacaggttaacagcgtcacaagaagaaccaagcagatcttggaactctt 1860
 Db 1801 CTGTGTCACTGTCTAAAGGTAAAGGTACCGGTCTAAAGAGAACACGCGATTTGGGAATCTTT 1860
 Oy 1861 gtcaagtcacatataatgagaggaagcagacatctaaagatggaagccttcgagtgat 1920
 Db 1861 GTCAAGTCCATTTAATGAGAGAGACACATCTTAAGATGGAAGCTTCGGGTGATGAT 1920
 Oy 1921 caactgataagcaataatgaggaatccctgtttggcagaagaacaaagaatgagcagtaa 1980
 Db 1921 CAACGATAGCAAGTAAATGAGAAATCCCTGTGGGCAAGACAAACAGATGCCATGGA 1980
 Oy 1981 accctaagaaggtctatgttctaactgaagcacaataagcagaatgatcagcttaattgt 2040
 Db 1981 ACCCTAAGAAGGTCTATGTCTACTAAGGCAATAAACGAGGAATATCATCAGTTTATTGT 2040
 Oy 2041 gcaaggaagaataagcaagtgcgaatgcaatgcaatgcaactggagaccccttgagactgag 2100
 Db 2041 GCAAGGAAGAATAAGCAAGTGCATGCAATGAGCTGAAGTCACTGGAGGCCCTTGAGACTGAG 2100
 Oy 2101 ctggccatctgaaacagcgttgatgataagagaagcaagaatctccactcccttaacgt 2160
 Db 2101 CTGGCCATCTGAAACAGCCTTGATGATAGAGAAACGAAGATTTCCCATTCCTCTACACT 2160
 Oy 2161 ggaatgagggagcgtgaacatcgccagcagaagaatgctgcccctcagtagaataagtg 2220
 Db 2161 GGAATGAGGGGCTTGATGATCCCCCAAGCAAAATGCTGCCCTCAGTAGAATATGGGT 2220
 Oy 2221 gagtcaaglaaatacagcgtgccctacagtgaatgacatgaccccaagatgacacgtlcat 2280
 Db 2221 GAGTCAAGTAAATACACACTGTCCTCCTACAGTATATGCCCAAGATGACACTGTCATT 2280
 Oy 2281 atagaagaatgacaggttccagatgttctcctcacatctctcagcagatccctccagc 2340
 Db 2281 ATAGAAGATGACAGGTTCCTGCTCTCCATCTCTGACACAGTCCCTCTCCACG 2340
 Oy 2341 tcccatgatagtgtgaggtttgtgacagcagatgctgtacttggtggccaagcgtcgaac 2400
 Db 2341 TCCCATGATGATGTGGGTTTGTGTGACGGCAGATGCTGTGACTTGGCCCAAGGCTCGAATC 2400
 Oy 2401 agtgaatcagcagactgctccttgagccagaatgltgatccagttctgtcttcaacga 2460
 Db 2401 AGTGAATCAGCCGACTGCTCTTGTGAGTCCAGATGTGATCCAGTCTCTGCTTTTAAACGA 2460
 Oy 2461 gaagaattgagacgtcagagatgtcagaanaagcacaagaacaatttcaatccag 2520
 Db 2461 GAAGAATTGAGCGTCAAGATGTGTGAGAAACGACAAACCAATTTCAATGCGCAAT 2520
 Oy 2521 caattgattctgcttaaacacgaanaatcaaaagcatgattagatagctacag 2580
 Db 2521 CAATTGATTCTGTTAAACACGAATAATCAAAACATGATTTAGGTATAGCTACGAG 2580

Oy 2581 actaaactcaatacagtgatgataaccagaagaagcaggtctcccaagcagatgtggtcct 2640
 Db 2581 ACTAAACTCAATAACAGTGTGATGATCAGAAAGCAGAGTTCTCCACGACGAGATGTGGGTCTCT 2640
 Oy 2641 tccctgggtcttaaaagatcaagctcgttgagagatctfcaacccgaatttcgaaggtg 2700
 Db 2641 TCCCTGGGTCTTAAGAAAGTCAAGCTCTGTGAGAGTCTGACAGCCGCAATTCGCCAGGTG 2700
 Oy 2701 acttgaatgagataatctcttcacatcgtccacggcccgagataatcagagcagga 2760
 Db 2701 ACTTTGAATGGGGATATTTCTTCCATCTGTCACGCCCGGATATATCAGAGCAGGGGA 2760
 Oy 2761 tgcgaatgagactcaagatgcgcacatgcacaatcttatgataaacccgcggtagatgat 2820
 Db 2761 TGCATGAGAGCTTCAGAGCTCCATCGACAAATCTTATGATAAACCCCGGTAGATGAT 2820
 Oy 2821 gatgataagcagatgagaccttggagaagaagcacagaagaagttcaagatcagggaga 2880
 Db 2821 GATGATGAAGGCAATGAGACCTTGGAAAGACACAGAAAGAAAGTTCAAGATCAGGGAGA 2880
 Oy 2881 gagctgtatccacagccagatgatacagcttcccaactccttgagagacaatgatagga 2940
 Db 2881 GAGCTGTATCCACAGCCAGTATGATCAGCTTCCCACTCTCTGAGAGACAAATGATGGA 2940
 Oy 2941 aaccaagagaaggtgataagaactgataaaaaaagataaaactggaagaagaagaag 3000
 Db 2941 AACCAAGAGAAAGGTGATGATGACTGTAGAAAAAGCATAAAACTGCAAAAGAAAAAGAG 3000
 Oy 3001 aaagataagataagagaagaagataaagaagccaagaagaagatgataagagcttg 3060
 Db 3001 AAAGATAGAGATGAAGAGAGATTAATGAAGCCMAAGAGGAATGCTGAAGGGCTTG 3060
 Oy 3061 ggaagacatgttcag 3075
 Db 3061 GGAGACATGTTCCAGG 3075

RESULT 5
 AK000761 2696 bp mRNA linear PRI 22-FEB-2000
 LOCUS Homo sapiens CDNA FLJ20754 fls, clone HEP02246.
 DEFINITION AK000761
 ACCESSION AK000761.1 GI:7021050
 VERSION o1igo capping; fls (full insert sequence).
 KEYWORDS Homo sapiens hepatoma cell_line:HePg2 CDNA to mRNA, clone_11b:HEP
 SOURCE clone:HEP02246.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
 Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
 Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
 Nakamura,Y., Isogai,T. and Sugano,S.
 TITLE NEDO human cdna sequencing project
 JOURNAL Unpublished (2000)
 REFERENCE 2 (bases 1 to 2696)
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
 Shibahara,T., Tanaka,T. and Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
 Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
 Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 COMMENT NEDO human cdna sequencing project supported by Ministry of
 International Trade and Industry of Japan; CDNA full insert
 sequencing; Research Association for Biotechnology; CDNA library
 construction; 5'- & 3'-end one pass sequencing; Depart of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).
 FEATURES
 source location/Qualifiers
 1..2696

4801 tggggagcgtggagacagagagtgctccctgcccgggggagagctgggcccattggccc 4860
1818 tggggagcgtggagacagagagtgctccctgcccgggggagagctgggcccattggccc 1877
4861 tgtgtctatcatagagagagagctaaagaataatctctctagagagagctaatggccc 4920
1878 tgtgtctatcatagagagagagctaaagaataatctctctagagagagctaatggccc 1937
4921 gtacatctagtaataatatttaattagttttgtctctacagctgtctagagagagcaca 4980
1938 gtacatctagtaataatatttaattagttttgtctctacagctgtctagagagagcaca 1997
4981 gaattggagacagagataaaccagacagcatttgaatctctctctacagctgttttcaagc 5040
1998 gaattggagacagagataaaccagacagcatttgaatctctctctacagctgttttcaagc 2057
5041 agagagcatttgaatc 5100
2058 agagagcatttgaatc 2117
5101 gctcctaagtagtctctgcatatgaatacaataaagagagagagcgtttacacacaca 5160
2118 gctcctaagtagtctctgcatatgaatacaataaagagagagagcgtttacacacaca 2177
5161 cagagagacatttgcgggtcccaacagcgggtggtgagagagagcctctcgcgtcagat 5220
2178 cagagagacatttgcgggtcccaacagcgggtggtgagagagagcctctcgcgtcagat 2237
5221 tagagtggt 5280
2238 tagagtggt 2297
5281 acatcctgattgtataacacatgctgggggtgtctaaagtgctgtgaatcccgatgtgaa 5340
2238 acatcctgattgtataacacatgctgggggtgtctaaagtgctgtgaatcccgatgtgaa 2357
5341 aaagctggaggttgaaagctcagacacacatgatttactttaaaaaacagaaaaaagaca 5400
2358 aaagctggaggttgaaagctcagacacacatgatttactttaaaaaacagaaaaaagaca 2417
5401 tgtatgagatgatttatt 5460
2418 tgtatgagatgatttatt 2477
5461 tagaataatgtgtccac 5520
2478 tagaataatgtgtccac 2537
5521 gtctcttaac 5580
2538 gtctcttaac 2597
5581 ggaatgaagtgagct 5640
2538 ggaatgaagtgagct 2657
5641 catatgcaactccttaataaagacatatttctctctgt 5679
2658 catatgcaactccttaataaagacatatttctctctgt 2696

RESULT 6
AF196186 3470 bp mRNA linear PRI 16-NOV-2001
LOCUS Homo sapiens atypical PKC isotype-specific interacting protein
DEFINITION short variant mRNA, complete cds.
ACCESSION AF196186
VERSION AF196186.1 GI:13491611
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 3470)
AUTHORS Fang, C.M. and Xu, Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted asip isoforms in human hepatocellular carcinomas
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE 21520080
PUBMED 11642408
REFERENCE 2 (bases 1 to 3470)
AUTHORS Fang, C. and Xu, Y.
TITLE Exon/Intron Structure and Splicing Variants of a Novel Human Polarity Gene, HASIP
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 3470)
Fang, C. and Xu, Y.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Laboratory of Molecular and Cellular Oncology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P.R. China

FEATURES
source
1..3470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
/note="HASIPa"
/codon_start=1
/product="atypical PKC isotype-specific interacting protein short variant"
/protein_id="AAK27892.1"
/db_xref="GI:13491612"
/translation="MKTVVCFGRFVVPCGDGHHKVESLIQAAVTRYKAIAKDPNY MIOVRLHGGGGLDLDLIDCDVADKDLVAVDEDDPHHGGGCTASSGTSPE IFSRLGNNVNSAOPVQATSEITSEVRLANMLPHRRSSDPLIGLSVSSSNP SSEPSRKNPRTSTTAFKONTAGTSPKTCDDKDENRSLPRTSMVSHQFORDNA RSSLSASHPWVGKLEKQEDQDETEEDNSRVEPVGADTGLEHLPNSLDDWKLVE VPNDGPIGLIHVPEFSARGRLGLVLRLEKGAHEHNLEFRENDCIVRIINDGLRN RFEQAOHMFQAMRTPLIMFHVYPAKQEYEOLOSSEKNNYSRSRPSQYIDNR SVNSAGLTQVORAPLNHPPOIDSHSLPHSAHSGRPPSAPAPAPNPVSTVSSG YNTRKIGKRNLQKKKEGEGFSTSRDVTIGSAPYVNNILPRGAJODGRKAC DRLIEVNGVDLVKQSEEVSLRSTKEGVSLVFNQEDAFHREINAEBSQMOIP KETRAEDDVLTPDGTREFFLEFVPLNDSGAGLVSVKGRNSENADIGLFEKSI INGGASDKDGLRLVNDOLIAVNGESLKTQNDQDAETLRNSGKRNENADIGLFEKSI RISKNEKLSGSPGPELPIETALDDEERRISHLVGIELDSERNALASIMG KYOISPTNMMODDVLIEDDLVLPPLHLSOSSSSSHDVEFTAAAGAPAKAIS DSADCSLSPVDVPAVAFORREGGRQSMSEKRTKORSQDOLDVFRSKSMGDIAD ETKLNTYDDQKAGRSRDVGRSLGKSSLSLESDTAYAEVTLNKDIFPHRRPRTIR GRGNESEFRAAIDKSIDKPAVDDDEGEETLEEDTESSRSGRSESVSTASDPSHLE ROMNNGNRKGDTRDKDKTGERKKRDKDKKAKKAKKMLKGLGDMFSLAKLPEK R"

BASE COUNT 1003 a 811 c 934 g 721 t 1 others
ORIGIN

Query Match 42.4%; Score 2414; DB 9; Length 3470;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 3064; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

1 atgaagtgacggt 60
288 atgaagtgacggt 347

61 atgaagtgacggt 120
348 atgaagtgacggt 407

121 gatccaactactgatacagtgacgcttggaacacatgagatgaggaatgactagac 180
408 gatccaactactgatacagtgacgcttggaacacatgagatgaggaatgactagac 467

181 ctgtatgacatcttctgtatgtatgacagacgataaagacagctgtgtgtgtgtgtgtgtgt 240


```
OY 2401 agtgaatcagccagctcctcttgatgacagatgttgatccagctcttgctttcaacga 2460
|||||
DB 2679 AGTGATGACGCGCAGTGCCTTTGAGTCCAGATGTTGATCCAGTCTTCTTTCAACGA 2738
|||||
OY 2461 gaagatttgagcgtcagagatgtcagaataaacgacacaaatcttcaatgcacat 2520
|||||
DB 2739 GAAGGATTTGGACGTCAGGATGTCAGAAAAACGACAAACCAATTTTCAATGCGCACT 2798
|||||
OY 2521 caattggatttcgttaaaacacgaataatcaaaacatgatgattagatagctgacag 2580
|||||
DB 2799 CAATTGGATTTCGTTAAACACGAAATCAAAACATGATTTAGCTTACTGACGAG 2858
|||||
OY 2581 actaaactcaatcagctgagatgcacgaagaagcttctccacagagatgtyggctct 2640
|||||
DB 2859 ACTAAACCAATACAGTGTGATGCCAGAAAGCAGGTTCTCCACAGCAGATGTGGGTCT 2918
|||||
OY 2641 tccctgggtctcfaaagaatcagctcgttggagagcttcacacgcgagttccgagtg 2700
|||||
DB 2919 TCCTTGCGGTCTGAAGAAAGCAAGCTCATTGGAGAGCTGTCAGACCGCAGTTGCCGAGTG 2978
|||||
OY 2701 acttgaaatggagatattcctcttcacatcgltcacagcgccgagataatcagaagcag 2760
|||||
DB 2979 ACTTTGAATGGGATATTCTTCCATGTCACGCGCGGATATATCAGAGCAGCGGA 3038
|||||
OY 2761 tgcgaatgagagcttcagagctgcacgcacaaactttatgataaacccggttagatgat 2820
|||||
DB 3039 TGCATATGAGAGCTTCAGAGCTGCCATCGCAATCTTATGATTAACCCCGGTAGATGAT 3098
|||||
OY 2821 gatgatgagagatgagagacttggaagaagacacagaagaagatcagaagcagagga 2880
|||||
DB 3099 GATGATGAAGGATGAGAGCTTGGAGAGACACGAAAGATTCAGATCAGGAGAGA 3158
|||||
OY 2881 gactgtgatccacagccagtgatcagccctccactctcttgagagacaaatgaaatga 2940
|||||
DB 3159 GAGTCTGATTCACAGCGCGATGATCAGCTTCCACTCTCTGAGAGCAATATGATGA 3218
|||||
OY 2941 aaccaaagagaagtgatgataagctatagaaaaagataaaactggaagaagaagaag 3000
|||||
DB 3219 AACCAAGAGAAAGGTGATGATGACTATGAAAAAGATTAACGTAAGAAAAAGAG 3278
|||||
OY 3001 aaagataagataaagagaagataaaatgaaagccaaagaaggaatgctgaaggcttg 3060
|||||
DB 3279 AAAGATGAGATTAAGGAGAGATTAATAATGAAGCCAAAGAGGAATGCTGAAGGCTTG 3338
|||||
OY 3061 ggaagacatgttcag 3074
|||||
DB 3339 GGAGACATGTTGAG 3352
|||||

RESULT 7
ALS91464 196374 bp DNA linear HTG 19-DEC-2001
LOCUS Homo sapiens chromosome 10 clone RP13-302D10, *** SEQUENCING IN
DEFINITION PROGRESS ***, 14 unordered pieces.
ACCESSION ALS91464
VERSION ALS91464.2 GI:17973978
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Dec 20, 2001 this sequence version replaced gi:14141513.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
```

```
----- Project Information
Center project name: b302D10
----- Summary Statistics
Sequencing program: XGAP; version 4.5
Sequencing vector: plasmid; L06755; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 189117 bases at least Q40
Consensus quality: 192136 bases at least Q30
Consensus quality: 193772 bases at least Q20
Insert size: 195074; sum-of-contigs
Insert size: 221513; 17.0% error; agarose-fp
Quality coverage: 4.57x in Q20 bases; sum-of-contigs Quality
Coverage: 4.33x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 40923: contig of 40923 bp in length
40924 41023: gap of 100 bp
41024 52424: contig of 11401 bp in length
52425 52524: gap of 100 bp
52525 56668: contig of 4144 bp in length
56669 56768: gap of 100 bp
56769 66064: contig of 9296 bp in length
66065 66164: gap of 100 bp
66165 74102: contig of 7938 bp in length
74103 74202: gap of 100 bp
74203 118025: contig of 43823 bp in length
118026 118125: gap of 100 bp
118126 120523: contig of 2398 bp in length
120524 120623: gap of 100 bp
120624 134346: contig of 13723 bp in length
134347 134446: gap of 100 bp
134447 143138: contig of 8692 bp in length
143139 143238: gap of 100 bp
143239 151820: contig of 8582 bp in length
151821 151920: gap of 100 bp
151921 179327: contig of 27407 bp in length
179328 179427: gap of 100 bp
179428 187106: contig of 7679 bp in length
187107 187206: gap of 100 bp
187207 193223: contig of 6017 bp in length
193224 193323: gap of 100 bp
193324 196374: contig of 3051 bp in length.
Location/Qualifiers
1. 196374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-302D10"
/clone_1fb="RP13-13.2"
1. 40923
/note="assembly:fragment:00704
fragment_chain:1"
41024. 52424
/note="assembly:fragment:02076
fragment_chain:1"
52525. 56668
/note="assembly:fragment:02444
fragment_chain:1"
56769. 66064
/note="assembly:fragment:00688
fragment_chain:1"
66165. 74102
/note="assembly:fragment:01900
fragment_chain:1"
74203. 118025
/note="assembly:fragment:02289
```

5	24114	1309	others
---	-------	------	--------

190006 GCGAAGCAGGAAATGCTACCTGCTGGAGCAGAACTACTCC 3736

Oy 3737 ctgggaaggtctccagaatgccaagaagaaaccccaagtctccaaactaaccaagatcca 190065
 Db 190066 ctggggaaggtctccagaatgccaagaagaaaccccaagtctccaaactaaccaagatcca 3796
 Oy 3797 ggaacggtcaccctgggaaggaatggtctcaagccaggtctcagtctgynaacttaagagc 190125
 Db 190126 ggaacggtcaccctgggaaggaatggtctcaagccaggtctcagtctgynaacttaagagc 3856
 Oy 3857 tccctcgccaaggaacaagaagcggaagagcagacaaatgaagaacagccctccctcggaag 190185
 Db 190186 tccctcgccaaggaacaagaagcggaagagcagacaaatgaagaacagccctccctcggaag 3916
 Oy 3917 ggcacaagaactatgactgactaagaagaagccaggaaccccaagttaagccctcccaag 190245
 Db 190246 ggcacagacacactgactgactaagaagaagccaggaaccccaagttaagccctcccaag 3976
 Oy 3977 ggccttcgcgaagaatgctgcacccctccctctcaggttcgaggttgaaagacttc 190305
 Db 190306 ggccttcgcgaagaatgctgcacccctccctctcaggttcgaggttgaaagacttc 4036
 Oy 4037 aagacttcbaagaagaagagccctctatctcctbgaacggaataaagaaagactcaag 190365
 Db 190366 aagacttcbaagaagaagagccctctatctcctbgaacggaataaagaaagactcaag 4096
 Oy 4097 tgcgcaataaagaacatctctatbgaagotgtaattttggagctttttaaact 190425
 Db 190426 tgcgcaataaagaacatctctatbgaagotgtaattttggagctttttaaact 4156
 Oy 4157 cgatgtgactatgagatattctctgttctgtatagctcccttaagcggtgaggaac 190485
 Db 190486 cgatgtgactatgagatattctctgttctgtatagctcccttaagcggtgaggaac 4216
 Oy 4217 gaatggaagacctaatgctcttgcaacatgctcaagtgctgttcaibgaagat 190545
 Db 190546 gaatggaagacctaatgctcttgcaacatgctcaagtgctgttcaibgaagat 4276
 Oy 190547 gaaatggaagacctaatgctcttgcaacatgctcaagtgctgttcaibgaagat 190605

[illegible]

OY 5357 gctcagcataccatgtattcttcttaaaacagaaaaaacatgatatgtctta 5416
 DB 191686 GCTCAGCATACCATGTATTCTTAAACAGAAAAACACATGTGATGATGTTA 191745
 OY 5417 tttttttttatggacatgtatttctgttgaactgtcttttagaagatgtcc 5476
 DB 191746 TTTTtttttttttttGGACATGTATTtTtGTGTACCTGTttTAGAAATGATGTCTCC 191805
 OY 5477 acacacgtaaccggtgtctctcttgcattctgttgcattgtcttcttcaatcagtg 5536
 DB 191806 ACACACGTAACCCGTTCTCTCTGCATTTCTGTGCATGTCTGTCTTAAATCAGCTG 191865
 OY 5537 cggcggtgtcttaagtgtgttaccagtgtaacgagtgacgttgatgagacgtgtcct 5596
 DB 191866 CGGGGtGTCTTAAGTGTGTTACAGTGTACCCGATGACCTTGATGACAGTGTGCTCT 191925
 OY 5597 ttccacagctcccttgaactgttgaagaacagcttctctgtacattgtcaactcctaa 5656
 DB 191926 TTTTCACAGCCTCCCTCCTGAGCTGTGACAAACAGCTTTCTGTACATATGCAACTCTTA 191985
 OY 5657 taaaaggcatattcttctcgtt 5679
 DB 191986 TAAAGGCAATTTCTTCTCTGT 192008

RESULT 8
 AL160409/c 106198 bp DNA linear PRI 18-JUL-2001
 LOCUS Human DNA sequence from clone RP11-406D17 on chromosome 10,
 DEFINITION complete sequence.
 ACCESSION AL160409
 VERSION AL160409.12 GI:14970794
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 106198)
 AUTHORS Sycamore, N.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT requests: clonerequest@sanger.ac.uk
 On Jul 19, 2001 this sequence version replaced gi:14625535.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT;
 Tr: TrEMBL; Wp: WormPep; Information on the WormPep
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
 RP11-406D17 is from the library RP11-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/Dacpac/home.htm>
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-406D17. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.

The true left end of clone RP13-348N17 is at 104199 in this
 sequence. The true right end of clone RP11-490D24 is at 2000 in
 this sequence.

FEATURES	source	Location/Qualifiers
repeat_region	1..106198	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="10" /clone="RP11-406D17" /clone_11b="RP11-11.2" 27..160
repeat_region	989..1039	/note="MIR repeat: matches 3..139 of consensus"
repeat_region	1232..1275	/note="MIR repeat: matches 93..144 of consensus"
repeat_region	1344..1651	/note="22 copies 2 mer ac 77% conserved"
repeat_region	4229..4606	/note="AlusX repeat: matches 1..302 of consensus"
repeat_region	4648..4815	/note="MIR17 repeat: matches 1..403 of consensus"
repeat_region	5215..5516	/note="FRAM repeat: matches 4..171 of consensus"
repeat_region	6161..6216	/note="AlusX repeat: matches 1..300 of consensus"
repeat_region	6217..6527	/note="MIR repeat: matches 47..104 of consensus"
repeat_region	6528..6637	/note="AlusX repeat: matches 1..295 of consensus"
repeat_region	6646..6950	/note="MIR repeat: matches 104..205 of consensus"
repeat_region	8061..8354	/note="AluDb repeat: matches 1..304 of consensus"
repeat_region	8443..8602	/note="AluDb repeat: matches 1..289 of consensus"
repeat_region	10005..10072	/note="AluDb/FRAM repeat: matches 152..308 of consensus"
repeat_region	10542..10849	/note="L2 repeat: matches 2673..2741 of consensus"
repeat_region	10851..10975	/note="MER57B repeat: matches 43..386 of consensus"
repeat_region	11451..11758	/note="FLAM C repeat: matches 15..133 of consensus"
repeat_region	12992..14459	/note="AluDb repeat: matches 1..312 of consensus"
repeat_region	14460..14550	/note="PTR5 repeat: matches 504..1874 of consensus"
repeat_region	14551..14771	/note="MER61E repeat: matches 377..466 of consensus"
repeat_region	14772..14923	/note="PTR5 repeat: matches 1874..8522 of consensus"
repeat_region	14930..15218	/note="AlusX repeat: matches 3..154 of consensus"
repeat_region	15223..15519	/note="AluDb repeat: matches 1..293 of consensus"
repeat_region	15661..15818	/note="Alusg repeat: matches 1..299 of consensus"
repeat_region	15868..15919	/note="MER5A repeat: matches 34..189 of consensus"
repeat_region	16133..16443	/note="26 copies 2 mer aa 73% conserved"
repeat_region	16599..16707	/note="AlusX repeat: matches 1..312 of consensus"
repeat_region	16712..16787	/note="L1M4 repeat: matches 7728..7834 of consensus"
repeat_region	16825..17092	/note="38 copies 2 mer ta 76% conserved"
repeat_region	17094..17254	/note="AluDb repeat: matches 1..292 of consensus"
repeat_region	17579..17889	/note="AluDb/FRAM repeat: matches 134..295 of consensus"
repeat_region	18807..19086	/note="Alusg repeat: matches 1..306 of consensus"


```

misc_feature /note="assembly_name:Contig14"
1196..2203
/note="assembly_name:Contig15"
2304..9462
/note="assembly_name:Contig55
9563..end:??"
clone_end:??"
/note="assembly_name:Contig56
23508..39871
/note="assembly_name:Contig57"
39972..68325
/note="assembly_name:Contig58"
68426..138688
/note="assembly_name:Contig59"
BASE COUNT 41112 a 30000 c 29735 g 37221 t 600 others
ORIGIN

Query Match
Best Local Similarity 34.38; Score 1952; DB 2: Length 138688;
Matches 2002; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3677 ggcaagcaggaagaatgcaagctgctcccaagactctgggagcagaactactccc 3736
|||||
Db 136248 GGCAGACGAGGAAAAATGCGACCTGCTCCAGAGCTCTGGAGCAGAACTACTGCC 136189
OY 3737 ctgggagagcttcagagtgccaaagagaccccaagctactccactcccaagctcca 3796
|||||
Db 136188 CTGGGGAGAGCTTCCAGAGCTGCAAGAAAGAACCCAGAGTACTCCAGTACCAAGGCTCCA 136129
OY 3797 ggaacggcgaactgggagagacatggtctcaagccaggtcatgtctggaactcaggagc 3856
|||||
Db 136128 GGAACGGCTACTGCTGGAGAGCATGGCTCAACGCCAGGCTCATGCTGGAAGACTCAGAGAC 136069
OY 3857 tcccttcgcaggaacagagcaggaagagcagatgaagaacagccctcccgagg 3916
|||||
Db 136068 TCCTTCGCGAGAAACAGAGCGGAGAGACACAGATGAAAGACACCTCTCTCCGAGG 136009
OY 3917 ggcgcagcaactatgactgcttaagaagctccagaccacattggccctcccaag 3976
|||||
Db 136008 GGGCCAGCAACTGTGACTGTATGAAGAACCCAGAGCCCAATTGAGCCCTCCCAAG 135949
OY 3977 ggccttcgcaggaagatgtgccccctccctccctcaggttcggaagctgaagacttc 4036
|||||
Db 135948 GGGCCCTTCCGCGAGAGATGTGCCCTCCCTCTCAGGTCTCGAGGCTGAACAGACTTC 135889
OY 4037 agactcctggaagagagagccctctattctcggagcagcagaataaagagagcttcag 4096
|||||
Db 135888 AGACTCCTGAGAAAGAGAGGCTCTTATTCTGAGCAGCAAGTAATAAGGATCTTCATG 135829
OY 4097 tcggcgaataaagaacatttctctatgaagactgtaatttgggaggtttttaaact 4156
|||||
Db 135828 TCGCGCAATTAAGACATTTTCTTATGAAGACTGTGATTGGGAGTTTAAACCT 135769
OY 4157 cgatggtactatgagatattctgtgtgtgtatcagtgaccttaagcgggtgaagcaca 4216
|||||
Db 135768 CGATGCTACTATGAGATTTCTGTGTGTGATCAAGTCTTAAACGCGGTAGCAAA 135709
OY 4217 gaatggaagagcctaattgcttctgcaactatgctcaagtgctcgttcaatgaagaat 4276
|||||
Db 135708 GAATGGAAGGCGCTTAATGTCTTGCCACATGTCTCAAGGTCTGTTATGTAAGAT 135649
OY 4277 ttcccaacctgtgacaactatgctgttggagtgttcataatgctcggccctcccaagt 4336
|||||
Db 135648 TTCCCAACCTGTGTACATATCTGTTTGAAGGTTCATATCTCTGCGCTCTCCACAGT 135589
OY 4337 accagaatctggccctctcagatgttcgagcggctgtgtgtaaatcccgagc 4396
|||||
Db 135588 ACAGAGATCTGGCCCTTACTACTATGAGTTGCCGCGCTTGGTTGAAATCCCTGCAAC 135529
OY 4397 acttgagtgagaacttcaactgaagtgaggatgacgtgagccctgttccctcta 4456
|||||

```

```

Db 135528 ACTTGCACTGACAAATTTACCTGAGAGTGAGAGATGACGTGCGCCCTGTTTCCCTCTA 135469
OY 4457 agtctcttagtatgagatgacatcttagtctctgtgtggaagaaagtggcgagacatc 135469
Db 135468 AGTTCTCTTAGCTATGAGATGACATCTTACTCTGTGTGAGAAAGTGGGCAACATAC 135409
OY 4517 accaaaaattgggcttctcgttactcacaagcagcagccatttgcacttctacac 4576
Db 135408 ACCAAAAATGGGCTTTCGTGACTTCACAGACACAGCCATTTGTCTGACTTGTCTACA 135349
OY 4577 ctgtgtttctctcttctccttccctcagctcttgtgaaggagagtcgtacatctac 135349
Db 135348 CTGTGTTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 135289
OY 4637 agaaagtaagccatgctcacaatggttggtcacacatgggggtccttttgaactgct 4696
Db 135288 AGAAGCTAAGCCATAGTCCAACTATGTTGGTCAACATGGGGGTCTCTTGTGAACGTGCC 135229
OY 4697 tatgactcaactacataaagtagatgactgctgctgcttatacatagcttgt 4756
Db 135228 TATGACTCAACTTACCAATTAAGTATGATGCTGCTGCTGCTTAAACATACGCTTGT 135169
OY 4757 cgttccttctctgacacgtgggttgatgaccacacagctctgtgtgggagctgggaga 4816
Db 135168 CGGTCTGTCTCTGACACGTGGGTTAGTACACACAGCTCTGTGTGGGACGCTGGGAGA 135109
OY 4817 caggaatggtctctcgcggggagagctggcctgcacttgccctgttctatcatgag 4876
Db 135108 CAGGAGTGGCTCTCGCGGGAGACCTGGCTGCTCATGTCCTGTGTCAATGAGG 135049
OY 4877 ggaagactaagaaaaattctcctaggaagagctatgagcccaatgactatgact 4936
Db 135048 GGAAGACTTAAAGAAATTTCTCTAGAGAGCTATGAGCCATGATCTCTGTAAT 134989
OY 4937 atttaattgatttcttctcgaagcttgcagagcttgcagagcagaagaggaagata 4996
Db 134988 ATTTTAATTATTTTCTTCTGACACTTGTCAAGAGGAGCAAGAAATGGGACAGAGATA 134929
OY 4997 aaccagacagctatttgaactgtctctcagagtttcaagtaagaggaattatgact 5056
Db 134928 AACCAAGACACTCAATTTGATCTGCTCTACGCTTAAAGTCAAGGCAAGTATGATCT 134869
OY 5057 tgltaatgcatcacaacactgcatgctgactgagatgacagctcctaaatgactct 5116
Db 134868 TGTCTAATGATCAACACACTGCAATGCTGACGTGGCGATGCCGCTCTAAGTATCT 134809
OY 5117 gccatgaacatataaagaacaaagaaagccgttacaatcacaagagaacatttcgg 5176
Db 134808 GCCATGAACATTAATAAGAAAGAAAGCCGTTACACATCACACAGAGAACATTTTCGG 134749
OY 5177 gtcccaagcgggtgtgagagagctcctcctcgtcagatattagagtggtgtgag 5236
Db 134748 GTCCCAACGCGGTGGGAGAGAGCTCACTCCGTCAGTATTAAAGTGTGTGTGG 134689
OY 5237 gctcgggagatcctcgggtcccatctcctcatttgcatttgcataatccctglatglaa 5296
Db 134688 GTCGCGGGGATCGGTGCTGCCATCTTCCTCATGTTCTTAAACATCTTATTTGTA 134629
OY 5297 accatgctgggggtgctaaatgctgctgaaatccgagatggaagagctgagatgtaa 5356
Db 134628 ACCATGCTGGGGGTCTAAAGTGCCTGTAATCCATGTGAAAAAAGCTGAGAGTAA 134569
OY 5357 gttcagcatacatgatttactttaaaacagaanaaagacatgattgataatgctta 5416
Db 134568 GCTCAGCATACCAATGATTACTTTAAAGAAAGAAAAAAGACATGTATGATATGCTTA 134509
OY 5417 ttttttttattgagacatgatttattgtgtgactgttggaaatgagtgatgctc 5476
Db 134508 TTTTATTTTATTTGCAATGTATTTTGTGTGACTGTGTTTGAATGATGATGTC 134449
OY 5477 acaacgtaccggtgtcctctcgtcattctgtgataagttctgttcttaataagct 5536
Db 134448 ACACAGATACCCGCTGCTCTCTGCAATTTCTGTGATGAGTCTGTCTTATACCTG 134389

```

QY 5537 cggcggttctaagtgtgttctacagctgtacgcgacgtacgttgatgtacagtggtctt 5596
|||||
Db 134388 CGCGGTCGTCTAAGTGTGTGTATACAGTGTACGCGAGTACCTGTGATGACAGTGGCTCT 134329
QY 5557 ttctcacagcctccctcagctgtgtgagaacgcttctctgtacatgcaactcctaa 5656
|||||
Db 134328 TTCTTCACAGCCCTCCCTGTGAGTGTAGAAACAGCTTCTCTGTACATATGCAACTCTTAA 134269
QY 5657 taaagcatttcttctctgtt 5679
|||||
Db 134268 TAAAGGCATATTCTCTCTGT 134246

RESULT 10
AK024668 2116 bp mRNA linear PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ21015 f1s, clone CAE05730, highly similar to
DEFINITION AF252293 Homo sapiens PAR3 (PAR3) mRNA.
ACCESSION AK024668
VERSION AK024668.1 GI:10437004
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens primary endothelial cells of human coronary artery
cDNA to mRNA, clone_id:CAE clone:CAE05730.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hiki,J.T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,K., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEO human cDNA sequencing project
TITLE Unpublished (2000)
JOURNAL 2 (bases 1 to 2116)
REFERENCE 2 (bases 1 to 2116)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center,
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@life.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction: 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source Location/Qualifiers
1..2116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="primary endothelial cells of human coronary
artery"
/clone="CAE05730"
/clone_id="CAE"
/note="Cloning vector pME18SFL3"
1..2116
/note="highly similar to AF252293 Homo sapiens PAR3 (PAR3)
mRNA"
BASE COUNT 651 a 476 c 571 g 418 t
ORIGIN

Query Match 33.4%; Score 1899; DB 9; Length 2116;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1899; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2227 gtaataacacagcttccctacagtgaaatgacccaagaatgacacgtctcatatagaa 2286
|||||
Db 200 GGTAAATACACAGCTGTCCCTACAGTGAATATGCCCAAGATGACACTGTCTATTATGAA 259

QY 2287 gatgacaggttgcacagtgcttccacacatctctctgacacgttcccttccagctccat 2346
|||||
Db 260 GATGACAGAGTGTGACAGTGTCTCTCCACATCTCTGTACAGAGTCTTTCAGCTCCAT 319
QY 2347 gatgatgtggggttctgacgcgacagatgctgtacttgggccaaggtctgaatgat 2406
|||||
Db 320 GATGATGTGGGGTGTGTGACGGCAGATGCTGTACTGTGGGCCAAGGCTGCATTCAGTAT 379
QY 2407 tcagccgacgtctcttgatgctcagatgttgatccagttcttcttcttcaacagagaaga 2466
|||||
Db 380 TCAGCCGACTGCTCTTGAATCCAGATGTGATCCAGTCTTCTTCAACAGAGAAGA 439
QY 2467 ttggaagctcagagtatgtcaagaaaacgcacaaagaacatttcaagatgcaatg 2526
|||||
Db 440 TTGTGACGTACAGATATGTACGAAAAACGACAAAGCAATTTTCAGATGCGCAGTCAATTG 439
QY 2527 gattcgttaaaaacacagaaaatcaaaaagcatalgatttaagtalagctgcagactaaa 2386
|||||
Db 500 GATTTCGTTAAAAACAGAAATCAAAAAGCATGGATTAGTATAGCTGACGAGACTAAA 559
QY 2587 ctcaataacagtgatgacccagaaagcaggtctctccacgacagatgtgggtcttccctg 2646
|||||
Db 560 CTCAATACAGTGGATGACCCAGAAAGCAGGTTCTCCACAGACAGATGTGGGTCTTCCCTG 619
QY 2647 ggtctgaagaaagtcacagctcgttggagagctgcagacccagcttgcgcagagtcattg 2706
|||||
Db 620 GCTCTGAAGAGTCAGAGCTGTGTGAGAGTCTGACAGCCCATTTGCCGAGGTGACTTTTG 679
QY 2707 aatgggataatcttcttccatcgttccacggcgcgcgataatcagaggaaggatgcaat 2766
|||||
Db 680 AATGGGATATATTCCTTCCATCCTCCACAGCCGCCGCGATATACAGAGGCAGAGGATGCAT 739
QY 2767 gaagagcttcagagctgcacatcgacaaactctatgatataacccgcggtatgatgat 2826
|||||
Db 740 GAGAGCTTCGAGCTCTCCATCGACAAATCTTATGATTAACCCCGGTGATGATGATGAT 799
QY 2827 gaagagatggaagccttggaaagaaacacagaaagaagtcacagatcagaaggaagagct 2886
|||||
Db 800 GAAGCATGAGAGCCTTGGAGAAAGACAGAAAGAAAGTTCAAGATCAGGAGACAGAGTCT 859
QY 2887 gttctcacagccagctgatacagcttccactctcttgagaagacaatgatgtaaaccaa 2946
|||||
Db 860 GATATCCACAGCCGATGATATCGCTTCCACTCTCTGGAGGACAAATGATGGAACCA 919
QY 2947 gaagaagtgataagactgataagaaaagaagtaaaactgtgaagaagaagaagaagaat 3006
|||||
Db 920 GACAAAGGTGATTAAGACTGATTAAGAAAAGATTAAGTAAGTAAGTAAGTAAGTAAGT 979
QY 3007 agagataaggaagaagataaaatgaaagccaagaaggaatgctgaaaggtcttggagac 3066
|||||
Db 980 AGAGATTAAGAGAAAGATTAATAATGAAGCCAAAGAGGAATGCTGAAGGGCTTGGAGAC 1039
QY 3067 atgttcaggttggcaaacatcgaaaagaatgataagatgtgagaacacgggtataataaa 3126
|||||
Db 1040 ATGTTTACAGTTTGGCAAACTCGAAAAGATGACAAAGTTGAGAAAACGGGTAAATAA 1099
QY 3127 atcacgaatccttcatcatcaagaagaagaagatacgaatgaaagcagagcagaggaag 3186
|||||
Db 1100 ATTCAGGAATCTTTCATTCATCAGAAAGAGAGAGATTCGAATGAAAGCAGGAGAGG 1159
QY 3187 attcaagccaaaactcgagaatattgaagaaacgaagctcgaagcgtgtgactatgtgaa 3246
|||||
Db 1160 ATTCAGGCCAAAACCTCGAGAAATTAAGGAACGACAGCTCGAAGCGTGTACTGTGAA 1219
QY 3247 attcaagatttcatcgcgacatttggctgtgatagatagtatgataatgtggggagttct 3306
|||||
Db 1220 ATTCAAGATTTTTCATGCGACATTTGGCTGTGATGATGAGTATATGTGGGGAGTTTCT 1279
QY 3307 tcttaagaaagttcactatgctcctcaacgctcagaagccctcagaagcccaagaagaagcatg 3366
|||||
Db 1280 TCTTATGAAGTTCATGTGCTCTCAACGCTTACAGCTCAGAGCCACGAGAAAGGCAATATG 1339

[illegible]

QY 4337 accgaagatctcgccctactcaatgagttgltccggcgttggtgtgaacatccctgacc 4396
|||||
Db 10128 ACCAGAAATCTGGCCCTACTCATGAGTTGTCGGCTTGGTTGAACATCCCTGACCC 10187
QY 4397 acttgcagagacaaatccaccggaagtgagagatgacgtagcgccctgtttccctctca 4456
|||||
Db 10188 ACTTGACAGTGAACAAATCACTGAAATGGAGATGACGTGCGCCCTGTGTTCTCCCTGTA 10247
QY 4457 agttctcttagctatgagatgacatcttagtctctgtgtgagagaaagtggcgacatac 4516
|||||
Db 10248 ACTTCTCTTAGCTATGSGGATGACATCTTAGTCTGTGTGAGAGAAAGTGGCGACATAC 10307
QY 4517 accaaaatggggccttctggttaactcaagacaaagccattgtctgtaacttctgtaaca 4576
|||||
Db 10308 ACCAAAAATGGGGCTTCTGCTACTTTCACAGCACAGCCATTGTGCTACTTGTTCATCA 10367
QY 4577 ctgtgtgtttctcttcttcttcttcttcaagctcttctgagaggaagtcgcatctcttacc 4636
|||||
Db 10368 CTGTGTGTTTCTCTTCTCTTCTCTGACGCTTGTGTGAGGGAGAGTGTGATCCTATTTAC 10427
QY 4637 agaagcctaagccatagtcacacatgttctgtgtcaacatgggggtcccttctgaactgctc 4696
|||||
Db 10428 AGAAGCTAAGCATAGTCCAAATGTTTGTGTGACCATGGGGTCTTGTGTAACGTCCCT 10487
QY 4697 tatgactcaacatccaaataagtgatgattccctgggtctgtgttatacatagcgttgtt 4756
|||||
Db 10488 TATGACTCAACATTAACAAATTAAGTATGATCTGTGCTGTGCTTTATCATACGCTTGT 10547
QY 4757 cggctcgtgtctcagacagtggtgtgagtcacacagcctctgtgtggaagcgttgagaa 4816
|||||
Db 10548 CGGTCTCTGTTCTGTACACGCTGGGTTGAGTACACACAGCTCTGTGTGGAGACGTGGAGAA 10607
QY 4817 caggagtggtctcgtccgggggaagcgtgggcctggtccattgtgcccctgtgtctatcatgag 4876
|||||
Db 10608 CAGGATGGGCTCTCGCCGGGGAGAGGTGGGCTGTGCATTTGGCCCTGTGCTATCATGAGAG 10667
QY 4877 ggaagagctaagaagaatctctctagaagagctcaatggccagttacatcctagtaatt 4936
|||||
Db 10668 GGAGAGCTAAGAAATTCCTCTAGAGAAAGCTCATGCCCCAGTACATCTCTAGTAAT 10727
QY 4937 atttaattagtttctgtctctcaagcgtgtctagaagaggcagaatgtggcagaagata 4996
|||||
Db 10728 ATTTTAATTAAGTTTGTGTCTGACAGCTGTGTGAGAAAGGCGACAGATGGGACAGAGATA 10787
QY 4997 aaccagacagtcatttgaatctgctctctctacagtttttcaagtcaagagcaattgtgct 5056
|||||
Db 10788 AACGAGACAGTATTTTATCTGCTCTAGCGTTTTCATAGTCAGAGGCATTTGATGCT 10847
QY 5057 tgtctaattgcatccacacacatgcatgtctgac-tggcgaatgcacagcctcctaagtatgc 5115
|||||
Db 10848 TGTCTAATGCAATCCACACACTGCTCATGTGATGGCGATGCCACGCTCTTAAGTATGTC 10907
QY 5116 tggccttgaaacataaagaagaagaagccgtttacacatcacacagaagaacatttgg 5175
|||||
Db 10908 TGCCATGAAACATTAAGCAAAAGCAAAAGCCGTTTACATCACACAGAGAAATTTTGG 10967
QY 5176 ggtcccaacagcggtgtgtgcaagagctcactctcggtcagttatagagttgtgtgtg 5235
|||||
Db 10968 GGTCCACACGCGGTGTGGCAGAGACTCTCGCGTCATATTAGAGTGTGTGTG 11027
QY 5236 ggtctcgaggatctcggtgtgtcccatcttctctcaattgttctgaacatccgtatgtga 5295
|||||
Db 11028 GGTCTCGGGGATCTCGGTGGCTCCCATCTTTCCTCATTTGTCTGAAACATCTGTATTGTA 11087
QY 5296 aacctgtctgggtgtgtctaaagtgtcgttgaaatcccgatgtgagaaagcgtgaggtgaa 5355
|||||
Db 11088 AACCATGGCTGGGGTGTCTAAATGTGCTGTAATCCGATGTGAAAGAAAGCTGAGAGTGAA 11147
QY 5356 agctcagcatcacatgtatcttcttaaaaaacagaaaaacacatgtgtgtgatgtct 5415
|||||
Db 11148 AGCTAGCATACCATGTATTATTCTTTAAAGAAAGAAAAAAGACATGTATGATATGTCT 11207

QY 5416 atttttttttattggcacatgtatatttgtgtgtgactgtttttagaagatgtagtgc 5475
|||||
Db 11208 ATTTTATTTTATTTGGCACATTTATTTTGTGTGACTGTTTGTGAAATGATGTGTCT 11267
QY 5476 cacacagtaacccgtctctctctctgcatctctgtgtgcatgagttctgttttaacagct 5535
|||||
Db 11268 CACACAGTACCCGCTGCTCTTCTGCAATTTCTGTGTGATGTTGTGTCTTAAATCCGT 11327
QY 5536 gcgcggtgtcttaagtgtgttaccagtgtaacgcagtgactgtgatatgagtgctc 5595
|||||
Db 11328 GCGCGGTGTCTAAGTGTGTGTACAGTGTACCGGAGTGTACCTTGATGTACAGTGTGCTC 11387
QY 5596 ttcttcaacagctccctcgtgagctgtgtgagaaagccttctctgttacaatgtaactccta 5655
|||||
Db 11388 TTTCTCACACCTTCCCTGTGAGTGTGAGAAAGCTTTCCTGTATATGCAATCTCTTA 11447
QY 5656 ataaagagcatatttctctctgt 5679
|||||
Db 11448 ATTAAGGCAATTTTCTCTCTGT 11471

RESULT 12

BD004905 2718 bp DNA linear PAT 31-JAN-2002

LOCUS BD004905 Novel genes encoding protein Kinase or protein phosphatase.

DEFINITION BD004905 Novel genes encoding protein Kinase or protein phosphatase.

ACCESSION BD004905.1 GI:18632866

VERSION BD004905.1 GI:18632866

KEYWORDS JP 03074935-T/2.

SOURCE Homo sapiens.

ORGANISM Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 2718)

Authors Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,

Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K., Otsuki,T.,

Funahashi,S., Sano,C. and Nezu,J.

Novel genes encoding protein Kinase or protein phosphatase

Patent: JP 03074935-T 2 30-JAN-2001;

HELIIX RESEARCH INSTITUTE,TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,

KOJI HAYASHI,KAORU SAITO,JUNICHI YAMAMOTO,SHIZUKO ISHII, OMOYASU

SUGIYAMA, AI WAKAMATSU,KEIICHI NAGAI,TETSUJI OTSUKI,SHINICHI

FUNAHASHI, HIKI SENO, JUNICHI NEZU

OS Homo sapiens (human)

PN JP 03074935-T/2

PD 30-JAN-2001

PF 28-JUL-2000 JP 200005061

PR 29-JUL-1999 JP 99P 248036,11-JAN-2000 JP 00P 118776 PR

02-MAY-2000 JP 00P 183767,09-JUN-2000 JP 00P 241899 PR

18-OCT-1999 US 60/159590,17-FEB-2000 US 60/183322 PT TOSHIO

OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI, PI KAORU SAITO,

PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,

PI KEIICHI NAGAI,TETSUJI OTSUKI,SHINICHI FUNAHASHI,CHIAKI SENO,

PI JUNICHI NEZU

PC C12N15/12,C12N9/12,C12N9/16,C12N1/15,C12N1/19,C12N1/21, PC

C12N5/10,C07K16/40,

PC C12Q1/68,C12Q1/42,C12Q1/48

CC

FE FT Key Location/Qualifiers

1..2718 Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 840 a 586 c 688 g 604 t

Query Match 29.8%; Score 1694; DB 6; Length 2718;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1844; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 374 ttgaggtcacaccttaagctctctgagcaaatgcctcttcaatgtgttgagcagtagtg 433

[illegible]

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2718)
AUTHORS Isogai, T. and Otsuki, T.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 299-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction;
5'- & 3'-end one pass sequencing and clone sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source
Location/Qualifiers
1..2718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OVARC1000945"
/tissue_type="ovary, tumor tissue"
/clone_id="OVARC1"
/note="Cloning vector: pME18SFL3"
33..2630
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55330.1"
/db_xref="GI:14042634"
/translation="MPLHVRSSDPALIGLSTYSDSNFSSEPSRKNPTRMTTAF
LKONTASPKACDKDENYRSLPRDTSNMGNORNDKLVESHPVGMKQKOE
ODEDGEDNSRVRPVGADHGLIHPNLSIDMDVAVKLVENDDGPIGLIHVPFARG
GRITGLVYKLEKCKGAHEHLFRENDCIVAINGDILNRNRFBOQHFRQAMRTPII
WFHVVPANKEQYEQLSQSEKNNTYSSEFSPDSYIDNRVNSAGLATVQAPRLNHP
PEQIDSHSRPLPSHAPSKPSPASAPQUNFSTVSGVYTKRIGKLTOLKKGTE
GLGFTISRDVTIGSAPYIKVNIILPRGAALODGRALKEVNGDYLNGSQEEV
VSLRSTKMEGTVALVYRQDAHPRELNPEPMOYPKETKAEDEDIVLPQSTR
FLFEVPLDSGAGLGVSVGNRSKEMHADLGFVSIINGKAASDGRIRVNDLI
AVNGESLIGKTNODAMETLIRKSMTEBGRKRMTOIYARRISKNEKLSRSPPEPEL
PIETALDDREKRISHLSYSGIEGDESFRNALISDGADSLSPVAMPDDVYIIE
DRLPLVDPHLSIDSSSSSHDDVGFVADATKAKA1ISDADSLSPVAMPDDVYIIE
EGFGRQDETKLNTVDQKAGSPRDVSPSLGKSSLSLEIQTAVAEVLTNGSDYFPI
RPRRIIRGRGNCSPFRAIDKSYDKPAVDDDGEMTLEEDTSSSRSGSEVSTAS
DQPSHSLEROMNGNDEKDKTRKKDKTKGKKKKRDKKKMAKKKMLKGLDMFS
LAKKPEKR"

BASE COUNT 840 a 586 c 688 g 604 t

ORIGIN

Query Match 29.8%; Score 1694; DB 9; Length 2718;
Best Local Similarity 99.8%; Pred. NO. 0;
Matches 1844; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 374 ttgaagtcacacacctcaagtccttcagcaaatatgctcttcattgctcgaagcaagtagt 433
|||||
Db 1 TTGAGGTCAACACCTTCACTCCTTCGAGCAAAATATGCCCTTCAATGTTGACGACCACTAGTG 60
QY 434 accagccttaattggtcctccacctctctgcagtgtatgaatttcctcccttgaagagc 493
|||||
Db 61 ACCAGCCTTAATTGGCTCTCCACTTCTGTCAGATAGTAATTTTCTCTGTAAGAC 120
QY 494 ctcaagaaaaatccacacagctgtgtcaacacacagctgctctcctcaagcaagacagt 553
|||||
Db 121 CTTCAAGAAAAATCCACACGCTGCTCAACACAGCTGGCTTCTTCAAGCACACTG 180
QY 554 ctggagagtcctctctctctctcgaagaaagatgaaaaatctacagaaagcctccgcgag 613
|||||
Db 181 CTGGAGTCTCTAAAGCCCTGCGACAGAGAAAGATGAAGAACTACAGAACCCCTCCGCGGG 240
QY 614 atactagtaactgtgtctaaccaattcagagagacaaatgctgcgtctctgtaagagc 673
|||||
Db 241 ATACTAGTAAGTGTCTTAACCAATTTTCAGAGAGCAATGCTGCTGCTCTGATGTCGA 300
QY 674 gtcaaccatgtggtggaagctggtggaagaaacagaaatgagatgagtaggagagag 733
|||||

Db 301 CTCACCCAAATGGTGGGCAAGTGGCGACAGACAACAAACAGATGAGATGGACAGAG 360
QY 734 aggataacagtcgtgtgtgaacctgtgtgaacatgtgtgacaggglttgagacatalacca 793
|||||
Db 361 AGGATTAACAGTGTGTGTAACCTGTTGACATGCTGACACGGGTTTGAGACATATACCA 420
QY 794 actttctctgatatgtatgttaagctcgtgaagtcctcccaagatgagggcctctgg 853
|||||
Db 421 ACTTTCTCTGTGATGTATGTAAGAACTGTGMACTCCCAACGATGAGGGCCTCTGG 480
QY 854 gaatcatgtatgaccttcaagtgctcgaagcgcgagaaacctgggggttatatgaataaac 913
|||||
Db 481 GATTCATATAGTGCCTTTCAGTGTGAGCGCGCAACCCCTGGGGTTATTAATAAAC 540
QY 914 gattgagaagaagtggtlaaagctgtacaatgaaatcttctctgtagaatgattgattg 973
|||||
Db 541 GATTGAGAAAGTGTGTAAGCTGAACATGAAATCTTTTGTGTGAAATGATGATGATTG 600
QY 974 taagattatgaatggcgacctcgaatagaaagatttgaacagacacataatcttc 1033
|||||
Db 601 TCAGGATTAATGATGGCGACCTTCGAAATGAAAGATTGTAACAGCAACATATGTTTC 660
QY 1034 gccaaagcctgcgttaacccatcattgtgtccatgtgttctcctgagaagaaataaagaac 1093
|||||
Db 661 GCCAAGCCATGCGTACACCCATCATTTGTTCCATGTGTTCTCGACCAATTAAGAGC 720
QY 1094 agtatgaacacatctccaaagtgtgaagaacacatctactatctcaacgcgttttaagccct 1153
|||||
Db 721 AGTATGAACACATGATCCCAAGTGAGAAAGCAATTAATTAACCCGTTTAAAGCCCTG 780
QY 1154 acaagcagtatattgcaacacaggaagtgtgaacagtgcaagggcttcaacgggtgcaagag 1213
|||||
Db 781 ACAGCCAGATATTGTAACAACAGAGGTGTGACAGTGACAGGCTTCACAGGTCACAGAG 840
QY 1214 caacccagcttgaaaccccgccctgagacagatagacctctcaatgaacacctcattagag 1273
|||||
Db 841 CACCCGCAATGAACCAACCGCTGAGACAGATACACTCTCACTAAGACTCTCATAGGC 900
QY 1274 caacacccctcgaggaaaaac 1333
|||||
Db 901 CACACCCCTCGGGAANAACCATCGCTCCACCGTCCAGGACCTCGAATGATTTAAGTA 960
QY 1334 cgaactgaagcagtggtgttataacacacacacacacacacacacacacacacacacacac 1393
|||||
Db 961 CGACTGTAGCAGTGTGTTAAACACCAAAAAAATAGGCAAGAGGCTTAATATCACACTTA 1020
QY 1394 agaaagtagtaagaagtttggattcagcatcctccagaagatgtaaaataatgagtgct 1453
|||||
Db 1021 AGAAAGTAGACGAAGGTTTGGATTGACATCTTCCAGAGATGTAACAAATAGTGAGCT 1080
QY 1454 cagctccaatctatgtgaaaaaatctctcccgaggggcgcgcaattcagatgagcgcagc 1513
|||||
Db 1081 CAGCTCCATATATGTAAGAAACATTTCTCCCGGGGGGGGCGCATTTACAGATGGCGGAC 1140
QY 1514 ttaagggcaggaacagacttataagagtgtaagtgagtagattagtgtggcaaatcccaag 1573
|||||
Db 1141 TTAAGGCAAGAACACACTTATAGAGTAATGAGATTAATTAGTGGCAAAATCCCAAG 1200
QY 1574 aggaagctgttcgctgtgtgagaagcacaagaatggaaggaactgtggaacctctggtct 1633
|||||
Db 1201 AGGAAGTGTTCGCTGCTTGAGAGACACCAAGATGGAAGAACTGTGAGCTTCTGTGCT 1260
QY 1634 ttgcgcagaagaacgctctccaccccaagggaactgaaatgcagaagccaaagccagatgcaga 1693
|||||
Db 1261 TTCGCAGAGAAAGCGCTTCCACCCAAAGGAAGTGAATGTAATGCAGACCAAGATGCAGA 1320
QY 1694 ttccaaagaagaacgaagaagatgaggaatatgtttcttacacctgtgtgacccaagg 1753
|||||
Db 1321 TTCCAAAAAGAAAGCAAGAGATGAGATATGTTCTTACACTGATGACACAGAGG 1380
QY 1754 aattctgacattgaagcccaacttaagatctcagagatctgaagcctctgtgtcagtg 1813
|||||
Db 1381 AATTTCTGACATTTGAAGTCCCACTTAGATTAACAGATCTCAGGCTTGTGTTGACGTG 1440

[illegible]

Db 661 TCTCTGAGTGGCCAGTACCCCAATGGTGGCAAGTGGCTGGGAACAAACAAAGAGATGAG 720
Qy 721 gatggagaagaagatacaacagtcgtgttgaaactgttgagacatgcagcgttg 780
Db 721 GATGGAGAGAAGAGATACAGTGTGTGTGAACCTGTTGGACATGCTACACGGGTTTG 780
Qy 781 gagcatataccaactctctctgtgatatgtataagtcgtgtaagtcctcccaagat 840
Db 781 GAGCATATACCCCAACTTCTGTGATGATAGTAAAGCTGTAAGAGTCCCAACGAT 840
Qy 841 ggaaggcctctggagatcatgtagtgcttcaatgctcggagcgcgagaaacctggg 900
Db 841 GGAGGGCCTCTGGGATCCATGTAGTGTCTTCAATGCTCGAGCGCGCAACCTGGGG 900
Qy 901 ttattagtaaaacgacttggaagaagtgtaagcttgacaagaatcttctctgag 960
Db 901 TTATTAGTAAACGATGTGGAGAAAGTGTGTAAGCTGAACATGAAATCTTTTCTGTGAG 960
Qy 961 aatgattcattgtcagaattaaatgagcgacctcgaataagaagattgaacaaga 1020
Db 961 AATGATTCATGTCTGAGATTAATGATGCGGACCTTCGAAATAGAGATTTGAACAAGCA 1020
Qy 1021 caacatgatttcgcaagacgaatgctgtaacccatcatcttgctcatgttgcttga 1080
Db 1021 CAACATATGTTTCGCAAGCCATGCGTACACCCATCATTTGCTTCATGTGTTCTCTGA 1080
Qy 1081 gcaataaagaagcaatatagaacaactatcccaagtgagaagaacaattactatcaag 1140
Db 1081 GCAATTAAGAGCAATATGAACAACTATCCCAAGTGAAGAAACAAATATACATTCAAC 1140
Qy 1141 cgttttagccttgacagcagatataatgaacaagaagtgagaaagtgagagcttca 1200
Db 1141 CGTTTACCCCTGACAGCCAGTATGTCACACAGAGAGTGAACAGTGCAGGCGTTTCA 1200
Qy 1201 acggcgagagagagaccccgacttgaaaccccgctgagcaatgactctcaactcaaa 1260
Db 1201 ACGGCGAGAGAGACCCCGACTGTAACACCGCTGAGCAATGACATCTCACTCAACAA 1260
Qy 1261 ctactcctagcgacacacccctcgggaaacaacacatccgcgcacgctcggcacctag 1320
Db 1261 CTACTCTATAGGCGACACCCCTCGGGAAACACCATCCGCTCCAGCCCTCGGACCTGAG 1320
Qy 1321 aatgtattatgaagcagtcgtaagcagtggtataaacaacaaaataagcagaagctt 1380
Db 1321 AATGATTTTATGACACCTGTAAGCAGTGTATTAACACCAAAAAAATAGCGAAGGCTT 1380
Qy 1381 aatatacagccttaagaagaagtgacagaagtggtgagattcagatcacttccagaatga 1440
Db 1381 AATATCCAGCTTAAGAAAGGTACAGAAAGTTTGGGATTCAGCATCACTCCAGAGATGTA 1440
Qy 1441 acaatagtgagtcagctccaactcattgtgaaacaactctcccccgggggagcgacatt 1500
Db 1441 ACAATAGTGGGTGTCAGCTCCATATGTGAAAAACATTTCTCCCGGGGGGGGCGCAT 1500
Qy 1501 cagatgagcagacttaaggcagaagacagactatagaagtaaatagtagtaagt 1560
Db 1501 CAGGATGGCCGCTTAAGGACAGACAGACTTAAAGAGTAAATGAGTAAATTTAACTTAACTG 1560
Qy 1561 ggaacaatcccaagagaagtggttgcgtgttgagaagaacaaagaatgagaagactgt 1620
Db 1561 GCAATATCCCAAGAGAAAGTTGTTGCTGTGAGAAAGCAACAAAGATGGAAGAACTGTG 1620
Qy 1621 agcctctgctcttcgcaagaagaagccttccacccaagaaggaaactga 1670
Db 1621 AGCCTTCTGCTTCTTCGCCAGGAAACGCTTCCACCCCAAGGAACTGAA 1670

RESULT 15
AF332593 3822 bp mRNA linear PRI 16-NOV-2001
LOCUS AF332593
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long
variant b mRNA, complete cds, alternatively spliced.

ACCESSION AF332593
VERSION AF332593.1 GI:14579310
KEYWORDS
SOURCE
ORGANISM human
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Fang, C.M. and Xu, Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
asip isoforms in human hepatocellular carcinomas
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE 2152080
PUBMED 11642408
REFERENCE 2 (bases 1 to 3822)
AUTHORS Fang, C. and Xu, Y.
TITLE Exon/Intron Structure And Splicing Variants of A Novel Human
Polarity Gene, hASIP
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 3822)
TITLE Direct Submission
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
Oncology, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
320 Yue-Yang Road, Shanghai 200031, P. R. China
LOCATION/Qualifiers
1..3822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..3822
/note="ASIPb; alternatively spliced; similar to GenBank
Accession Number AF196185"

CDS

Accession Number AF196185
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein long variant b"
/protein_id="AAK69193.1"
/db_xref="GI:14579311"
/translation="MKVNYCFRTVYVPCGSHMKVPSLIGQAVRYKATAPNY
WQVRLRGDSGLDLDLIDVADKDRVAVVDEDDPHGGGTASSTGTGSP
ITGSELGTNNVSAFQPYQATSEIEVTPSLRNMPDHLVRSDDPLIGTSSQNF
SSEPSRKNPRTWSTAGLKTONTAGSPCTCRKDKDENTRSRPTSMNSNFORDA
VNSLSASHPVWKLEKQEDGDEEDNSRVEPVHADGTGLEHLPNFDLMVKLVE
VNDGCPGLGIVHVPSSARGSPRLGLVYKLEKGAHEHLEPNDICIVINDGLRN
RFEQAOHMFROAMRPTLIWEHVVAANKEVOELOSSEKNVYSRSRSPDOYIDNR
SVNSGHTVQARPLNHPPEIDISHSLRPHSAHSGKPPSPAPASPOVETSTYSSG
YNTKTRKRLNLTQKGTGEGSTSRVTTGGSPATVKNILPRGALIDGRLAG
DLILVNGVDLVKSGQEVVSLRSTKMGVSLVFRQEDAHPRELKAEDEDIVLT
PDGTRFLTFEVLNDGSGAGLVSVKGNRSKENADLGIPIVKSIIINGCAEDGRRL
VNDOLIVNGESLKTNDQDAEMLRSRSTGKNGMOLIVARIISGNCLEKSPGS
PPGPELPIETFLDDBRRRTSHLSYSGTEIDLSGKSNMALSLRMKYOLSPVNNPOD
DVITLEDRLPLVPLPHLSQSSSSSHDDVGPYTAAGTAAKAIISDASCSLSPVDP
VLAQREGRQIADETKLNITVDQAGSPSDVSGSLAKTSLSLSESLQIVAEVTL
NDIDPFHRRPRRIIGRGCESEFRAIDKSYDKPAVDDDEMETLEEDTESSSGR
ESVSTAQDPSSLSLRONMNGEKDGTDRKDKTGKKEKMDKEDKAKKMLK
GIGDMFRIQAKTREFREROARREDYAIODFFRTGCDDELWYGVSSYEGSMALNAR
POSPREGHMDALVIOVKKPRNSKSPVDSNSTSPNHRIORLROEFGQAASLDEVE
DDBRTVSEFQPPNAPATOSGRHSVSVQOROREROERESQOARQVSLPDRS
KNAVSSQDSWEQNTSPGEGFQSAKENPRTSYQSRNGLGGHGNVAVLETDLL
ROEQRKEQMKKPPSEBSPSYDSYKRVQDSVAPPKGPFROVDVPPSPQVARRNL
QPEKGRFYS"

BASE COUNT 1114 a 885 c 1019 g 804 t

Query Match 29.4%; Score 1670; DB 9; Length 3822;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgaaagtaacggtgtgcttcggaagcaccgggtgtgtgtcgtgcgtgagggagccac 60
|||||

```

Db      1 ATGAAGTGAACCGTGTGCTTCGAGAGGAGCCGGGTGTGTCGTCCGTGCGGAGCGGCAC 60
QY      61 atgaagtttcagctcatccagacgggtgacccgctacccggagggccatcgcaag 120
Db      61 ATGAAGTWTTCAGCCTCATCCAGAGCGGTGACCCGCTACCGGAAGGCCATCGCAAG 120
QY      121 gacccaactactgatacagaagtgatcgcttggaacatggaaatggaaatctaac 180
Db      121 GATCCAAACTACTGATACAGTACAGTGCATCGCTGGAAACATGGAGATGAGAACTATAC 180
QY      181 ctgagtcactcttctgtatgtagcagagcagaataagacaagactgtgtagcagtgat 240
Db      181 CTTGATGACATCTTCTGTGATGTAGAGAGAGATTAAGACAGACTGTGTAGAGTGTGAT 240
QY      241 gagcaggaatccacatccagcggaggtgtagcaccagtgccaggtccacgggtaccagagc 300
Db      241 GAGAGAGATCCACATCCAGGAGGTGAGCACCAGTGCACATTCACGGGGTACCCAGAGC 300
QY      301 ccagaagatttgtagagcttgcaaccaaatgtctcagcttccagcttaccag 360
Db      301 CCAGAGATTTTGTGTAGTAGAGCTTGCAACCAATGTCTCAGCCTTCAGCCTTACCA 360
QY      361 gcaacaagtgaattgagtcacacactcagcttccttgagcaaatatgtccttcatgt 420
Db      361 GCACCAAGTGAATTTGAGGTACACCTTCACGTCTTCGAGCAAAATGTGCTTTATGTT 420
QY      421 cgaacgaatgagaccagcctcctaattgacctcactctgtcagtgatgaatatt 480
Db      421 CCAGCAGTACTGACCCAGCTCTAATTTGGCTTCCTCACCTTCTGTGTAGGATGAAATTT 480
QY      481 tccctgagagccttcaagaaatccacacgctgttcaacaacagcgtggttctc 540
Db      481 TCCCTGAGAGCCCTTCACAGAAATCCCAACGCTGTGTAAACAGCTGGCTTCTC 540
QY      541 aagcagacactgctgagagtcctaaacactgagcaggaagaaagatgaaatcaga 600
Db      541 AAGCAGAACCTGCTGGAGTCTCTTAAACCTGCGACAGAAAGAAATGAAACCTACAA 600
QY      601 agcctccgggagactacgaatcgtgctaaccaattagaagagacaatgctcgtcg 660
Db      601 AGCCTCCCGGAGTACATGTAAGTACCTGTACCAATTTTACGAGAACATGCTCGCTG 660
QY      661 tctctgagtgccagtcacccaaatgtgtggaagtgagcaggaagaaagaaatgag 720
Db      661 TCTCTGAGTGCACGTCAACCCATGTGTGGCAAGTGTGCGAGAAACAGAACAGATGAG 720
QY      721 gatggacagaagaaggaatacagtcgttgaacctgttgacatgacagggcttg 780
Db      721 GATGGACAGAGAGATTAACAGTGTGTGAACCTGTTGACATGCTGACACCGGCTTG 780
QY      781 gagcatatcccaacttctctgagatgatatgtgtaagctcgtagaagtcaccaagat 840
Db      781 GAGCATATACCCAACTTTCTCTGATGATATNGTAAAGCTGTGAAAGTCCCAACGAT 840
QY      841 ggggggctctgggaatcagtagtgcttcaagtgctgaggcggcagaacccctgggg 900
Db      841 GGGGGGCTCTGGGAATCATGTAGTGTCTTTCAGTGTGAGGGCGCAAGACCCCTGGGG 900
QY      901 ttattagtaaaacgattggaagaaagtgtgaagctggaacatgaataatctttctgtag 960
Db      901 TTATTAGTAAACATTTGGAAGAGGTGTAAGCTGAACATGAATAATCTTTTCTGTAG 960
QY      961 aatgattcattgtcaggaataatgtagcagcctcgaaatagaagaatttgaacaagca 1020
Db      961 AATGATTTCATTGTCAAGATTAATGATGGCGACTTCGAAATAGAAATTTGAACAAGCA 1020
QY      1021 caacatagtttcgcaagccatgcttacaccatcatttggttcacatgtgttctgtga 1080
Db      1021 CAACATATGTTTGCCAAGCCATGTGTACACCCATCAATTTGTTCCATGTGTCTTCTGCA 1080
QY      1081 gcaataaagagcagtatgaacaactacccaagtgagaagaacaattactatctcaagc 1140
Db      1081 GCAAATTAAGAGCAGTATGAACAACCTATCCCAAGTGAAGAACTACTATTCAAGC 1140

```

Search completed: July 24, 2002, 14:33:32
Job time: 19019 sec

```

QY      1141 cgtttagccctgacagccagtatattgacaacaggagtgtaacagtgcagggtctaac 1200
Db      1141 CGTTTACCCCTGACAGCGATATATTTAGACAAAGAGTGTATTAACAGATGAGGCTTAC 1200
QY      1201 aaggtgacagaagaccccgactgaaacacccgcttgagcagatagactcactcaaga 1260
Db      1201 AAGGTGACAGAGACCCCGACTGAACACCCGCTGAGCAGATTAAGTCTCACTCAAGA 1260
QY      1261 ctactatagcgcacacccctccggaaacacacatcgcctccagcctggaacctcag 1320
Db      1261 CTACTCTATAGCGCACACCCCTCGGAAACACCATCCGCTCCAGCTTGSCACTCTCAG 1320
QY      1321 aatgatttagtagcagctgtaagcagtggtatataacaccaaataatgaagaagagct 1380
Db      1321 AATGATTTATGTAGTACAGCTGTAAAGCAGTGTATTAACACCAAAAATTAAGCAAGAGCTT 1380
QY      1381 aatattcagcttaagaaggtacagaaggtttggatcagatcactccagagatgta 1440
Db      1381 AATATCCAGCTTAAAGAAAGGTACAGAAAGTTTGGATTACGATCATCTCCAGAGATGA 1440
QY      1441 acaataggctgctcagctcaactctatgtgaaacatctcccccggggggggccatt 1500
Db      1441 ACAATAGGTGCTCAGCTCCAACTATGTGAATAACATTCCTCCCGGGGGGGCCATT 1500
QY      1501 caggatggccgacttaaggcaggaagacagacttatagaagtaaatggatagatagtg 1560
Db      1501 CAGGATGGCCGACTTAAAGCAGAGACAGACTTATAGAGTAAATVGATGATTTAGTG 1560
QY      1561 ggcgaatcccaagagaagttgttcgctgttgagaagccaagatlggaagaactgtg 1620
Db      1561 GGCGAATCCCAAGAGAGATGTTTGTGCTGTGAAGACCAAGATGGAAGAACTGTG 1620
QY      1621 agccttctgtcttccgcaaggaagccttccacccaagggaactgaa 1670
Db      1621 AGCCTTCTGTCTTTCGCAAGAGAGCGCTTCCACCAGGAAGAACTGAA 1670

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:33:29 ; Search time 21.33 Seconds
(without alignments)
6108.633 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356
Sequence: 1 MKVTVCGRFRVVPVPCGDGH.....SOVARLNRLQTPKGRPPYS 1356

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	6.5	1337	2 T13948	atypical protein k
2	11	0.8	1464	2 T13716	bazooka gene prote
3	11	0.8	2450	2 S71625	protein-tyrosine-p
4	9	0.7	1693	2 S76086	beta transducin-11
5	8	0.6	88	2 T15027	hypothetical prote
6	8	0.6	107	2 T39293	hypothetical prote
7	8	0.6	128	2 C86179	hypothetical prote
8	8	0.6	153	1 D71357	probable cation-ac
9	8	0.6	219	2 E84117	hypothetical prote
10	8	0.6	283	2 C64379	hypothetical prote
11	8	0.6	317	2 T27179	hypothetical prote
12	8	0.6	318	2 T24509	hypothetical prote
13	8	0.6	348	2 I49262	hypothetical prote
14	8	0.6	395	2 AD0380	cyclin cdk inhibit
15	8	0.6	450	2 G01158	multidrug efflux p
16	8	0.6	475	2 T38597	tyrosine kinase ac
17	8	0.6	592	2 S43570	zinc finger protei
18	8	0.6	618	2 AB0186	C0585.6 protein (c
19	8	0.6	623	2 H87127	probable exported
20	8	0.6	720	2 A45436	conserved hypotet
21	8	0.6	721	2 D87753	synapse-associated
22	8	0.6	723	2 I38412	protein C43E11.6 l
23	8	0.6	724	2 JH0800	receptor tyrosine
24	8	0.6	737	2 S44862	postsynaptic densi
25	8	0.6	767	2 T09599	R05D3.2 protein -
26	8	0.6	852	2 T10811	postsynaptic densi
27	8	0.6	852	2 A72343	channel associated
28	8	0.6	870	2 C01974	hypothetical prote
29	8	0.6	882	2 I38912	receptor tyrosine

30	8	0.6	890	1 A53743	protein-tyrosine k
31	8	0.6	904	2 I38757	homolog of Drosoph
32	8	0.6	911	2 I56552	synapse-associated
33	8	0.6	926	2 I38756	homolog of Drosoph
34	8	0.6	1081	1 A42399	isolucine--trNA 1
35	8	0.6	1149	2 T27567	hypothetical prote
36	8	0.6	1281	2 T00346	hypothetical prote
37	8	0.6	1333	2 E84601	probable retroelem
38	8	0.6	1551	2 AB2410	MD-repeat protein
39	7	0.5	58	2 S75890	hypothetical prote
40	7	0.5	79	2 D69309	conserved hypotet
41	7	0.5	84	2 A13640	hypothetical prote
42	7	0.5	96	2 D64695	type I restriction
43	7	0.5	99	2 T35967	hypothetical prote
44	7	0.5	104	2 T28823	hypothetical prote
45	7	0.5	104	2 H72553	hypothetical prote

ALIGNMENTS

RESULT 1
T13948
atypical protein kinase C isotype-specific interacting protein ASIP - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T13948
R:Zumd, Y.; Hirose, T.; Tamai, Y.; Hirai, S.; Nagashima, Y.; Fujimoto, T.; Tabuse, Y.
J. Cell Biol. 143, 95-106, 1998
A:Title: An atypical PKC directly associates and colocalizes at the epithelial tight
A:Reference number: Z17827; MUID:98437350
A:Accession: T13948
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1337 <120>
A:Cross-references: EMBL:AB005549; NID:93868777; PIDN:BA034216.1; PID:93868778
C:Genetics:
A:Gene: asbp

Query Match 6.5%: Score 88; DB 2; Length 1337;
Best Local Similarity 100.0%; Pred. No. 8.9e-82;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 566 PKETRAEDIEDIVLPDGTREPTFEVPLNDGSGAGLVGKNSKRNHADLGIFVKSII 625
|||||
DB 566 PKETRAEDIEDIVLPDGTREPTFEVPLNDGSGAGLVGKNSKRNHADLGIFVKSII 625
OY 626 NGGAASKDGRRLVNDQLIANGESLAK 653
|||||
DB 626 NGGAASKDGRRLVNDQLIANGESLAK 653
RESULT 2
T13716
bazooka gene protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13716
R:Kuchinke, U.; Grave, F.; Knust, E.
submitted to the EMBL Data Library, November 1998
A:Description: Control of spindle orientation in Drosophila by the Par-3-related PDZ-
A:Reference number: Z17708
A:Accession: T13716
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1464 <KUC>
A:Cross-references: EMBL:AJ130871; NID:e1363519; PID:e1363520; PIDN:CA10224.1
C:Genetics:
A:Gene: bazooka
A:Cross-references: FlyBase:FBgn0000163
A:Map position: X

Query Match
Best Local Similarity 0.8%; Score 11; DB 2; Length 1464;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 885 KKSSLESLGT 895
DB 1081 KKSSLESLGT 1091

RESULT 3

S71625
protein-tyrosine-phosphatase (RC 3.1.3.48) RIP - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C:Date: 27-Nov-1997 #sequence,revision 12-Dec-1997 #text_change 20-Jun-2000
R:Accession: S71625, S67987
R:Chida, D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Matanabe, T.
FEMS Lett. 358, 233-239, 1995
A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very early stage of mouse embryonic development
A:Reference number: S71625; MUID:95145716
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHT>
A:Cross-references: EMBL:DB3966; NID:q1232103; PIDN:BA12158.1; PID:q1232104
R:Wolf, B.B.; Brown, M.D.
FEBS Lett. 376, 177-180, 1995
A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound src tyrosine kinase
A:Reference number: S67987; MUID:96105375
A:Accession: S67987
A:Molecule type: protein
A:Residues: 1098-1102 <WOL>
A:Experimental source: submaxillary glands
C:Genetics:
A:Gene: Ptpn13
A:Map position: 5
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;
E:566-860/Domain: protein 4.1 membrane-binding domain homology
F:1089-1165/Domain: GLGF domain homology <GLG1>
F:1361-1437/Domain: GLGF domain homology <GLG2>
F:1495-1574/Domain: GLGF domain homology <GLG3>
F:1769-1840/Domain: GLGF domain homology <GLG4>
F:1863-1937/Domain: GLGF domain homology <GLG5>
F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>
F:2374/Active site: Cys (phosphotyrosine intermediate) #status predicted
F:2380/Binding site: substrate phosphate (Arg) #status predicted

Query Match
Best Local Similarity 0.8%; Score 11; DB 2; Length 2450;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 502 DGRKAGDRLI 512
DB 1800 DGRKAGDRLI 1810

RESULT 4

S76086
beta transducin-like protein, 190K - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein sl10163
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence,revision 25-Apr-1997 #text_change 26-May-2000
R:Keneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A:Reference number: S74322; MUID:97061201

A:Accession: S76086
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1693 <KAN>
A:Cross-references: EMBL:DB3999; GB:AB001339; NID:q1001396; PIDN:BA10064.1; PID:d101
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
F:1051-1084/Domain: WD repeat homology <WD01>
F:1092-1125/Domain: WD repeat homology <WD02>
F:1133-1166/Domain: WD repeat homology <WD03>
F:1174-1207/Domain: WD repeat homology <WD04>
F:1256-1289/Domain: WD repeat homology <WD05>
F:1297-1330/Domain: WD repeat homology <WD06>
F:1338-1371/Domain: WD repeat homology <WD07>
F:1420-1453/Domain: WD repeat homology <WD08>
F:1461-1494/Domain: WD repeat homology <WD09>
F:1502-1535/Domain: WD repeat homology <WD10>
F:1584-1617/Domain: WD repeat homology <WD11>
F:1625-1658/Domain: WD repeat homology <WD12>

Query Match
Best Local Similarity 0.7%; Score 9; DB 2; Length 1693;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 381 RFPSPSOXYI 389
DB 1431 RFPSPSOXYI 1439

RESULT 5

T15027
hypothetical protein Y1112 - Yersinia pestis plasmid pMT1
C:Species: Yersinia pestis
C:Date: 20-Sep-1999 #sequence,revision 20-Sep-1999 #text_change 20-Sep-1999
R:Lindler, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.
Infect. Immun. 66, 5731-5742, 1998
A:Title: Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid
A:Reference number: Z18268; MUID:99043898
A:Accession: T15027
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-88 <LIN>
A:Cross-references: EMBL:AF074611; NID:q3883003; PID:q3883110; PIDN:AAC82770.1
C:Genetics:
A:Gene: Y1112
A:Genome: plasmid pMT1

Query Match
Best Local Similarity 0.6%; Score 8; DB 2; Length 88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 KSIINGCA 629
DB 3 KSIINGCA 10

RESULT 6

T39293
hypothetical protein SPBC1105.16c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence,revision 03-Dec-1999 #text_change 03-Dec-1999
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Ransperger, U.; Bothe, G.; Pohl, A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21841
A:Accession: T39293
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-107 <MCD>
A:Cross-references: EMBL:AL096851; PIDN:CAB50979.1; GSPDB:GN00067; SPDB:SPBC1105.16c

A:Experimental source: strain 972h-; cosmid c1105
 C:Genetics:
 A:Gene: SPDB:SPC1105.16c
 A:Map position: 2

Query Match 0.6%; Score 8; DB 2; Length 107;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 163 EEPKRNK 170
 DB 77 EEPKRNK 84

RESULT 7

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86179

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 Jensen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86179

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-128 <STO>

A:Cross-references: GB:AE005172; NID:g2494117; PIDN:AAB80626.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 0.6%; Score 8; DB 2; Length 128;
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 284 LGIHVPF 291
 DB 53 LGIHVPF 60

RESULT 8
 D71357
 probable cation-activated repressor protein (tror) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000

C:Accession: D71357

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utechtack, T.; McDo
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: D71357

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-153 <COL>

A:Cross-references: GB:AE001201; GB:AE000520; NID:g3322431; PIDN:AAC65157.1; PID:g332243

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0167

C:Superfamily: conserved hypothetical protein M0568

Query Match 0.6%; Score 8; DB 1; Length 153;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 303 VKRLEKGG 310
 DB 48 VKRLEKGG 55

RESULT 9

hypothetical protein BH3741 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: E84117

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E84117

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-219 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807460.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3741

Query Match 0.6%; Score 8; DB 2; Length 219;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 507 AGDRLIEY 514
 DB 41 AGDRLIEY 48

RESULT 10

hypothetical protein M0635 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: C64379

R:Balt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
 ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: C64379

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-283 <BU>

A:Cross-references: GB:U67511; GB:L77117; NID:g2826300; PIDN:AAB98633.1; PID:g1591347

C:Genetics:

A:Map position: F0R564790-565641

A:Start codon: GTC

Query Match 0.6%; Score 8; DB 2; Length 283;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 999 KKKDRKE 1006
 DB 245 KKKDRKE 252

RESULT 11

hypothetical protein Y54G11A.10 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T27179
 R;Mallis, J.
 Submitted to the EMBL Data Library, December 1998
 A;Reference number: Z20322
 A;Accession: T27179
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-317 <WIL>
 A;Cross-references: EMBL:AL034488; NID:e1359895; PIDN:CAA22459.1; CESP:Y54G11A.10
 C;Genetics:
 A;Gene: CESP:Y54G11A.10
 A;Introns: 55/3; 108/3; 175/3; 228/2; 253/3

Query Match 0.6%; Score 8; DB 2; Length 317;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 640 DDLIANG 647
 |||||||
 Db 245 DDLIANG 252

RESULT 12
 T24509
 hypothetical protein T05C12.8 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T24509
 R;Burton, J.
 Submitted to the EMBL Data Library, October 1995
 A;Reference number: Z19901
 A;Accession: T24509
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-318 <WIL>
 A;Cross-references: EMBL:Z66500; PIDN:CAA91309.1; GSPDB:GN00020; CESP:T05C12.8
 A;Experimental source: clone T05C12
 C;Genetics:
 A;Gene: CESP:T05C12.8
 A;Map position: 2
 A;Introns: 26/3; 65/3; 103/3; 204/3; 245/3

Query Match 0.6%; Score 8; DB 2; Length 318;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1302 PSEGPSN 1309
 |||||||
 Db 261 PSEGPSN 268

RESULT 13
 149262
 cyclin cdk inhibitor p57 - mouse
 N;Alternate names: CDI p57; cyclin-cyclin-dependent kinase inhibitor p57; p57-kip2
 C;Species: Mus musculus (house mouse)
 C;Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 05-Nov-1999
 C;Accession: 149262
 R;Lee, M.H.; Reynolds, I.; Massague, J.
 Genes Dev. 9, 639-649, 1995
 A;Title: Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure
 A;Reference number: A56463; MUID:95247027
 A;Accession: 149262
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-348 <RFS>
 A;Cross-references: EMBL:U20553; NID:g794295; PIDN:AAC52186.1; PID:g794296
 C;Keywords: cell cycle control

Query Match 0.6%; Score 8; DB 2; Length 348;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 431 PPSAPASA 438
 |||||||
 Db 139 PPSAPASA 146

RESULT 14
 AD0380
 multidrug efflux protein [imported] - Yersinia pestis (strain C092)
 C;Species: Yersinia pestis
 C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C;Accession: AD0380
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tiltball, R.W.; Holden, M.T.G.; Prentice, M.
 deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
 Nature 413, 523-527, 2001
 A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A;Reference number: AB0001; MUID:21470413; PMID:11586360
 A;Accession: AD0380
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-395 <KUR>
 A;Cross-references: GB:AL590842; PIDN:CAC92367.1; PID:g15981070; GSPDB:GN00175
 C;Genetics:
 A;Gene: acra

Query Match 0.6%; Score 8; DB 2; Length 395;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 505 LKAGDRLI 512
 |||||||
 Db 354 LKAGDRLI 361

RESULT 15
 G01158
 tyrosine kinase activator protein 1 (TKA-1) - human
 C;Species: Homo sapiens (man)
 C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 16-Dec-1998
 C;Accession: G01158
 R;Closek, T.
 Submitted to the EMBL Data Library, July 1995
 A;Reference number: H00587
 A;Accession: G01158
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-450 <CTO>
 A;Cross-references: EMBL:Z50150; NID:g1246762
 C;Superfamily: GIGF domain homology
 F;16-86/Domain: GIGF domain homology <GLI>

Query Match 0.6%; Score 8; DB 2; Length 450;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 509 DRLIEVNG 516
 |||||||
 Db 193 DRLIEVNG 200

Search completed: July 24, 2002, 14:36:07
 Job time: 158 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 06:40:28 ; Search time 6663.82/Seconds
(without alignments)
17865.288 Million cell updates/sec

Title: US-09-757-781-20
Perfect score: 5689
Sequence: 1 atgaagtgaccgtgtgtctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Match Length DB ID Description

1	5636.8	99.1	5958	9	AF196185	AF196185 Homo sapi
2	4069.4	71.5	4071	9	AF467002	AF467002 Homo sapi
3	4041.4	71.0	4062	9	AF467003	AF467003 Homo sapi
4	3837.4	67.5	3960	9	AF467004	AF467004 Homo sapi
5	3763.4	66.2	3933	9	AF467006	AF467006 Homo sapi
6	3529.8	62.0	3822	9	AF332593	AF332593 Homo sapi
7	3486.2	61.3	3801	9	AF252293	AF252293 Homo sapi
8	3347.4	58.8	3735	9	AF467005	AF467005 Homo sapi
9	3056.4	53.7	5500	10	AB005549	AB005549 Rattus no
10	3050.6	53.6	3470	9	AF196186	AF196186 Homo sapi
11	2952.2	51.9	4167	10	AY026057	AY026057 Mus muscu
12	2772.6	48.7	2967	9	AF332592	AF332592 Homo sapi
13	2677.4	47.1	2696	9	AK000761	AK000761 Homo sapi
14	2476.8	43.5	2718	6	BD004905	BD004905 Novel gen
15	2476.8	43.5	2718	6	AK027735	AK027735 Homo sapi
16	2079.4	36.6	2116	9	AK024668	AK024668 Homo sapi
17	2005.2	35.2	196374	2	AL591464	AL591464 Homo sapi
18	2003.6	35.2	106198	9	AL160409	AL160409 Human DNA
19	2003.6	35.2	138688	2	AC025823	AC025823 Homo sapi
20	1982.2	34.8	22279	2	U82210	U82210 Homo sapien
21	1868.8	32.8	2004	9	AF177228	AF177228 Homo sapi
22	1695	29.8	1772	9	AK025892	AK025892 Homo sapi
23	1546.8	27.2	1592	9	BC011711	BC011711 Homo sapi
24	516.2	9.1	597	6	AX187060	AX187060 Sequence
25	482.4	8.5	540	6	AX321805	AX321805 Sequence
26	427	7.5	427	6	AX332939	AX332939 Sequence
27	415.4	7.3	425	6	AX072873	AX072873 Sequence
28	387	6.8	104357	9	AL392123	AL392123 Human DNA
29	357.4	6.3	504	6	AX185634	AX185634 Sequence
30	282.6	5.0	305	6	AX187694	AX187694 Sequence
31	282.6	5.0	309	6	AX186627	AX186627 Sequence
32	282.6	5.0	168213	2	AC103175	AC103175 Rattus no
33	273.2	4.8	4176	9	AF428250	AF428250 Homo sapi
34	271.6	4.8	339	6	AX184822	AX184822 Sequence
35	261.6	4.6	2656	9	AK056157	AK056157 Homo sapi
36	256.8	4.5	2068	9	AK057965	AK057965 Homo sapi
37	253.8	4.5	400	6	AX071341	AX071341 Sequence
38	246.8	4.3	70087	2	AC105981	AC105981 Mus muscu
39	245.4	4.3	165146	2	AC011644	AC011644 Homo sapi
40	245.4	4.3	167163	2	AC013705	AC013705 Homo sapi
41	245.4	4.3	169337	2	AL356462	AL356462 Homo sapi
42	245.4	4.3	193131	9	AL390766	AL390766 Human DNA
43	182.8	3.2	109816	9	AL360233	AL360233 Human DNA
44	182.8	3.2	138322	2	AC012241	AC012241 Homo sapi
45	182.8	3.2	196259	2	AL392171	AL392171 Homo sapi

ALIGNMENTS

RESULT 1
AF196185
LOCUS Homo sapiens atypical PKC isotype-specific interacting protein long
DEFINITION Variant mRNA, complete cds.
ACCESSION AF196185.1 GI:13491609
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Fang,C.M. and Xu,Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
asip isoforms in human hepatocellular carcinomas
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE 21520080
PUBMED 11642408
REFERENCE 2 (bases 1 to 5958)
AUTHORS Fang,C. and Xu,Y.
TITLE Exon/Intron Structure and Splicing Variants of a Novel Human
Polarity Gene, hAsip

DATE-SEP

Dh 1668 AATATCCAGCTTAAAGAAAGGTACAGAGGTTGGATTTCAGCATACCTTCCAGAGATGTA 1727
Qy 1441 acaataggtgctcagctcccaatctatg tgaaaaacatctcccccggggggcagccalt 1500
Dh 1728 ACAATAGGTGGCTGACGTCCATCTATGTGAATAAACATTTCTCCCGGGGGGGCCGCAATT 1787
Qy 1501 caggaatggtcgaaacttaagcagagagacagactatagaagttaaatlygagtagaattagtg 1560
Dh 1788 CAGATGTGCCCACTTAAAGCAGAGACAGACTTATAGAGTAAATGAGTATGATTAGTG 1847
Qy 1561 ggcacatcccaaggaggaattgtttcgtctgtttgagaagcaccacaatgagaagaaatttg 1620
Dh 1848 GGCATATCCCAAGAGAGAGATTGTTTCCGTGTTGAGAACACCAAGATGGAGAAAGTGGTG 1907
Qy 1621 agcctctgtctcttcgcaggaagacgccttccacccaagggaacttgaatgacagaca 1680
Dh 1908 AGCCTTGCTGCTTTCGCCAGGAAGACGCTTCCACCCAGGAACTGATTCAGAGCA 1967
Qy 1681 agccagatgacatctccaaaagaagcagaagatgagaatattgttcttaacct 1740
Dh 1968 AGCCAGATGCAAGATTCCAAAGAAAGAGAGAGATGAGATATGTTCTTACACT 2027
Qy 1741 gatggcaccaggaattcttcaactttgaagttcccaatattgattcaggatctcaggc 1800
Dh 2028 GATGGCACCGAGGAATTTCTGACATTTGAAGTCCCATTAATGATTCAGAGATCTCAGGC 2087
Qy 1801 ctgtgtgtcagtgltcaaaagtaaccggtcaaaagaagacacagcagatttggaaactt 1860
Dh 2088 CTTGGTGTCACTGTCAAAAGTAACCGGTCAAAAGAGAACACGCAAGATTGGGAATCTTT 2147
Qy 1861 gtcaagttcatattatgagggagcagcatctaaagatlygaaggtctcggtgtaagt 1920
Dh 2148 GTCAAGTTCATATTATTAAGAGAGAGACACATTAAGATGAAAGCTTCGGGTGATGAT 2207
Qy 1921 caactgatagagtaaatlygagaaatccctgtttggcagaagacaaacagaatlygaa 1980
Dh 2208 CAACTGATAGCAATGAATGAGAAATCCCTGTTGGCAAGACAAACAGATGCCATGGAA 2267
Qy 1981 accctaaagaaggtctatgtctacttgaagcaataaacgaggaatgatccagctta 2040
Dh 2268 ACCCTAAGAAAGTGTATGTCTACTGTAAGCAATAAAGAGAAATGATCAGATTATTGTT 2327
Qy 2041 gcaagggagaataagcaaggtgcaatgagctgaagtaacctgaggagccccctgagcttag 2100
Dh 2328 GCAAGGAAATAGCAATGTGCAATGAGCTGAAGTCACTGGAGGCCCCCTGTGAGACTGAG 2387
Qy 2101 ctggccatctgaacagcgttgatgataagagaacgaagaattcccatccctctacagt 2160
Dh 2388 CTGCCCATTTGAACAGCTTGGATGATAGAGACGAAAGAAATTTCCCATTTCCCTTACAGT 2447
Qy 2161 ggaattgaggggcttgaatcagccccagcagaatactgccccctcagtagaataagtg 2220
Dh 2448 GGGATTGAGGGGCTTGATGAAATCGGCCAGAGAAATGCTGCCCTCAGTAGAGTAATG--- 2504
Qy 2221 gagtcaggttaataaccagcgtctccctcagtgaaatgagccccagaatgacagcttcat 2280
Dh 2505 -----GGTAAATACCACTGTCCCTACAGTGAATATGCCCAAGATGACACTGTCAAT 2558
Qy 2281 atagaagaatgacaggttgcaggtgtctctcaacatctcttgaccagctccctccagc 2340
Dh 2559 ATAGAAGATGACAGGTTCCAGTGTCTCTCCACATCTCTCTGACCAAGTCCCTCTTCCAGC 2618
Qy 2341 tcccatgatagtatgaggttctgtgacggcagatgcttgtaacttggccaaggtcgaatc 2400
Dh 2619 TCCCATGATGATGTGGGTTTGTGACGGCAGATGCTGTGACTTGGCCCAAGGCTCAATC 2678
Qy 2401 agtgaattcagcagcagctgtctcttgaagtcagatgcttgaatccaggtctctgtcttcaaga 2460
Dh 2679 AGTGAATTCAGCCGACTGTCTTGTGAGTCCAGATGTTGATCCAGTCTTCTTTTAAAGCA 2738
Qy 2461 gaaggaatttgacgtcagagatgctcagaaaaacgacaaagcaatttcagatgacagt 2520
Dh 2739 GAAGATTGGACGTGACAGATGTGTGCAAAAAACGCAAAAGCAATTTTTCAGATGCCAGT 2798
Qy 2521 caattggatttctttaaacaacagaaatcaaaaagcatgatttagtataagctgacag 2580
Dh 2799 CAATTGATTGCTGTAAACACAGAAATCAAAAGATGATTAGTATGAGTGCAGAG 2858
Qy 2581 actaaactcaatacagtgatgatacgaagaagcaggtctcccaagcagatgttggtcct 2640
Dh 2859 ACTAAACTCAATACAGTGTGATACCGAAGAAAGCAGGTTCTCCAGCAGAGATGTGGGTCTCT 2918
Qy 2641 tccctgggtcttgaagaagtcagagctcggttggagagctctcagaacgcgaattcccgagtg 2700
Dh 2919 TCCCTGGGTCTTAAGAGAGTCAAGCTTATGAGAGTCTCAGACGCCAATTTGCCAGAGTG 2978
Qy 2701 acttgaatlyggataatctcttccatctgtccacggcggcggtaatactcagagcagggga 2760
Dh 2979 ACTTGAATGGGATATTTCTTTCCATCTCCACGGCCCGGATATATCAGAGCAGGGGA 3038
Qy 2761 tgcgaatgagagcttcagagctcgcacgcgaataacttataagataaccgcggtatgat 2820
Dh 3039 TGCATGAGAGCTTCAGACCTCCCATCGACAAATCTTATGATAAACCCCGGTAGATGAT 3098
Qy 2821 gatgatgaaggcatgagagaccttggagaagaacagagaagaatcagaatccaggaga 2880
Dh 3099 GATGATGAAGGCTAGAGACTTTGAGAGACACAGAAAGAAATTCAGATCAGGGAGA 3158
Qy 2881 gagctgtatccacagcagtgatcagccttcccaactctctgtgagagacaatlygaatgga 2940
Dh 3159 GAGTCTGTATCCACAGCCAGTGTACGCTTCCCACTCTGTGAGAGACAAATGATGGA 3218
Qy 2941 aaccaagagaagtgatgataagctgataagaaaaagataaanaactggaagaagaagaag 3000
Dh 3219 AACCAAGAGAAAGGTGTATGAGCTGTATGAAAAAAGATTAAGATGGAAGAAAGAAAG 3278
Qy 3001 aagaatagagataagagaagaagataaataatgaaagccaaagaaggaatgcttgaaggcttg 3060
Dh 3279 AAGATAGAGATAGAGAGAGAGATTAATATGAAAGCCAAAGAGGAGATGCTGAGGGCTTG 3338
Qy 3061 ggaagacatgtcaaggttctgcaacaacatcgaanaagatlygaanaacaggttaaa 3120
Dh 3339 GGAGACATGTTGAGTTTGGCAAAACATCGMAAAGATGACAAAGATTGAAAAACGGGTAA 3398
Qy 3121 ataaaaatcagaagatccttcatcagaagaaggagagatacgaatgagaagcag 3180
Dh 3399 ATAAAAATACAGGAATCCCTTACATCAGAAAGAGACAGATACGAATGAAGCAGGAGAG 3458
Qy 3181 gagagattcagaaccaaactcgaagaatttaggaacagcagctgagaagcgttgactat 3240
Dh 3459 GAGAGATTCAAGCCCAAACTCGAGAAATTTAGGAGACGACAACTCGAGAGCTGACTAT 3518
Qy 3241 gctgaattcagaatttcatcagacatttggctgtgagatgagttaaatgtatggggga 3300
Dh 3519 GCTGAATTCAGAGATTTTCATCGAGACATTTGGCTGTGATGATGATGATGATGATG 3578
Qy 3301 gttcttcttaagaaggttcacatgctcacaagctagaagctcagaagcccaagagaag 3360
Dh 3579 GTTTCTTTTATGAAGGTTTCATGCTCAACGCTAGACCTCAGAGCCACAGAGAAAGG 3638
Qy 3361 catatgatgagcttgttatgcccagtcagaagaacgcgggaattccaaacccctaact 3420
Dh 3639 CATATGATGATGCTTTTGTATGCTCCCAAGTCAAGAAACCCCGGAATTTCCAAACCTTACCT 3698
Qy 3421 gtagacagtaacagatccaactcctaagcaatcatalgacgatacagcgttgaagcaaga 3480
Dh 3699 GTAGACAGTAAACAGATCAACTCTACATCATGATCGATACAGGCTGTGAGGCAAGAA 3758
Qy 3481 ttccagcaagcaagaagaatgaagaatgtagaagaatcgtcggcgagactatagtttgag 3540
Dh 3759 TTTTCAGCAAGCAAGCAAGATGAAGATGTAAGAGATCGTCGGCGGACCTATAGTTTGAAG 3818
Qy 3541 caacccttgccgaacagcagcggcgagcagcagagagcgagcaactcgggtgtccgtggag 3600
Dh 3819 CAACCTTGCCCAACGACGCGCCGCGAGACGACAGCGGCGACACTCGGTGCTCGAG 3878

QY	3601	gtgcagatgacgagcgcaagcgaggaagagcgagagctcccaagagcccaagcgcaag	3660
Db	3879	gtcgaagtgtgaagcgacagcgacgagagagcgacagagctcccaagagcccaagcgcaag	3660
QY	3661	tacagctctctgcctccgccaagagaggaataatgccaagctcagctccccaagctctgg	3938
Db	3939	tacagctctctgcctccgccaagagaggaataatgccaagctcagctccccaagctctgg	3938
QY	3721	gacacgaataactcccttgggaaagcttccaaagtgcacaagagagaaacccagtactcc	3780
Db	3999	gacacgaataactcccttgggaaagcttccaaagtgcacaagagagaaacccagtactcc	3780
QY	3781	agctacccaagctcccaagaaagagctccctggagagacatgagcttaacgccaaggtcatg	4058
Db	4059	agctacccaagctcccaagaaagagctccctggagagacatgagcttaacgccaaggtcatg	4058
QY	3841	ctggaaactcggagcgctcttcgccaagacaaagcggaagagagcagatgaaagaag	3900
Db	4119	ctggaaactcggagcgctcttcgccaagacaaagcggaagagagcagatgaaagaag	3900
QY	3901	cagctctctccgagaggcccaacacatagactcgtataaagaalcccaagccccagt	4178
Db	4179	cagctctctccgagaggcccaacacatagactcgtataaagaalcccaagccccagt	4178
QY	3961	taagccctcccaagggcgccctccggaagatgtgcgcccccctccctcgaagttgcg	4020
Db	4239	taagccctcccaagggcgccctccggaagatgtgcgcccccctccctcgaagttgcg	4020
QY	4021	aggctgaacagactcgaagctccctcgaagaagagagcgcccttatctcgaacgcaaa	4080
Db	4299	aggctgaacagactcgaagctccctcgaagaagagagcgcccttatctcgaacgcaaa	4080
QY	4081	taacgagcttcctcagtcgcgcacaaataaagacatttccatagaagacttatttggg	4140
Db	4359	taacgagcttcctcagtcgcgcacaaataaagacatttccatagaagacttatttggg	4140
QY	4141	agctttttaaaccctcgatgagtcacatgtagagatcttggttggtgatacagtgccctt	4418
Db	4419	agctttttaaaccctcgatgagtcacatgtagagatcttggttggtgatacagtgccctt	4418
QY	4201	aagcggtgtgaagcaagaataatgaaagcgcttaatgtcttgcacataagctcaaatgtgc	4260
Db	4479	aagcggtgtgaagcaagaataatgaaagcgcttaatgtcttgcacataagctcaaatgtgc	4260
QY	4261	gtttcattgaaagatttcccaaccctgagacaacaatcgtttgaagtgltcaatgctc	4538
Db	4539	gtttcattgaaagatttcccaaccctgagacaacaatcgtttgaagtgltcaatgctc	4538
QY	4321	tgcgccctccacaagacacgaagatctcggccctcactcactgaagtgttcgagcggtgtt	4380
Db	4599	tgcgccctccacaagacacgaagatctcggccctcactcactgaagtgtgttcgagcggtgtt	4380
QY	4381	gtaacatccctgcacacacttgagtgagaaacttaactgaagtggagagatgaagtcgagc	4440
Db	4659	gtaacatccctgcacacacttgagtgagaaacttaactgaagtggagagatgaagtcgagc	4440
QY	4441	ccgtgttccctccaaattctcttgcataatggatagacatctagctctctcgttggagga	4500
Db	4719	ccgtgttccctccaaattctcttgcataatggatagacatctagctctctcgttggagga	4500
QY	4501	aaagtggcgacatacacaataaattggggcgcttcgtaactcaagacagacatgtg	4778
Db	4779	aaagtggcgacatacacaataaattggggcgcttcgtaactcaagacagacatgtg	4778
QY	4561	tcgtactctgtcacatcagtggttctctcttcccttcccttcccttcttggagagagag	4620
Db	4839	tcgtactctgtcacatcagtggttctctcttcccttcccttcccttcttggagagagag	4620
QY	4621	tcgatacctcattataagaagctcaagacatagtccaacatgtttgltcaaccaatggaggt	4680
Db	4899	tcgatacctcattataagaagctcaagacatagtccaacatgtttgltcaaccaatggaggt	4680
QY	4681	cccttctgaactgccttatgactacaatatcaacaataagtgatgtcctgtctcggtt	4740

[illegible]

RESULT	2
AF467002	
LOCUS	
DEFINITION	4071 bp mRNA linear PRI 07-FEB-2002
	Homo sapiens partitioning-defective 3 protein splice variant a (PAND3) mRNA, complete cds, alternatively spliced.

ACCESSION AF467002
VERSION AF467002.1 GI:18568343
KEYWORDS
SOURCE human.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 4071)
AUTHORS Gao,L., Macara,I.G. and Joberty,G.
TITLE Multiple splice variants of PAR3 and of a novel related gene, PAR3L, produce functionally different proteins
JOURNLT 2 (bases 1 to 4071)
REFERENCE Gao,L., Macara,I.G. and Joberty,G.
AUTHORS Direct Submission
JOURNLT Submitted (10-JAN-2002) Center for Cell Signaling, University of Virginia, PO Box 800577, Charlottesville, VA 22908, USA
FEATURES
source location/Qualifiers
1..4071 /organism="Homo sapiens"
gene /db_xref="taxon:9606"
1..4071 /gene="PAR3"
/note="PAR3"
1..4071 /gene="PAR3"
/note="PAR3"
CDS /gene="PAR3"
/note="PAR3; multi-PDZ domain containing protein; alternatively spliced"
/codon_start=1
/product="partitioning-defective 3 protein splice variant a"
/protein_id="AAL76042.1"
/db_xref="GI:18568344"
/translation="MKYVCFGRVVVPCGDGHMKVSLIQAVTRYKAIAKDPNY MIOYRLHGGGGILLDDILCDVADDDRDRLVAYVEDDDPHGGGTSASSGTGSPS IFGEIETNNYSAROPYOATSEIETVTPVLAAMPDLHYRSDPALIDLTSTSVSNR SREPSKRNPTRMSTTAAFLKQNTAGSKTCDKRDENYRELPRSTWMSNOFQDNA RSLLSAPHVGMLEKQEDGTEEDNSVEPVGHADTGLEHLPNSLIDMVLVE VPNDGGPGIHVVPFSAARGRTGLVLRLEKGAHEHNFRENDCLVRINDGDLN RNFEOAOFERQAMKPTPIIMFVVPANKEQEQLSQSEKNYSSRFSPOQYIDNR SPSAGLHTVORAPRLNHPPOIDSHSLPASHSPGSPAPAPAPQNVSTVSG LNTKIGKRLNIOLEKTEGEGISITSDVYIGSAPLYVNIILPRGAIDGRLKAG DRLLENGVDIVKGSQBEVVSILSTRKEGVSLVFOQDAFHRREINAEBSQOIT KETRAEDDILVTPDGTRELFEPVPLNDSGAVSVKGRSKENADLGIYKSI INGASAKDGLRVNDOLIAVNGESLIGKTQDAETRLRSMTEGNRGMIOQLVAR RISKCNELKSGSPPELPJETALDDEERISHSLYIGIGLDESFRNALSTRIMG ESGYQLSPYVNMPODDVITIEDRLPYLPHLSQSSSSHDGVTADATWAKA AISDSADCSLPVDVYLAFOREGRGSMSEKRTKOPSDASQLODFVTKRKSMDIG IADPRKINRYNDOKAGSPSRDVGSLCKKSSLSLEIOTPAVAYTLNDIDIPHRPR IIRGRCNESFRRAIDKSYDKPAVDDDEGMEULEEDTEESSRGSRESVSTASDPH SLERQMNQNGKGTDKDKTKTEKKRDKDEKDKKARKGMLKGLGDMFRGKHR KDDKIEKTKGIKIGESFTSEERIIMKOEORIKQKTEFERROREDVYEIOPFHR TFCDDDELMTGVSSEYEGSMALNAPQSPREGIMMDALYAYOVKKRPNKSPVDNRS TFSNHDRIORLROFOQKODEVEDRRRTYSPEDQMNAPATPGBEFOGAKENPRYSY ROROEERSSQOAROYSLSLPROGRKNSVSYDSQWENYSPGBEFOGAKENPRYSY OGSRNGYLGCGFNARVYLETQELRLDQRRKBOOMKQPSSEGSNTDSTYKRVDP S YAPKGPEDVPPSPSOVARLNRLQTEKGRPEFS"
misc_feature 814..1074 /gene="PAR3"
/note="Region: PDZ domain"
misc_feature 1378..1638 /gene="PAR3"
/note="Region: PDZ domain"
misc_feature 1765..2028 /gene="PAR3"
/note="Region: PDZ domain"
BASE COUNT 1220 a 923 c 1079 g 849 t
ORIGIN

Query Match 71.5%; Score 4069.4; DB 9; Length 4071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 atgaagtgaccgtgtgtcttcggaacgaccgggtgtgtcgtccgttgcgggaacgacc 60
DB 1 ATGAAGTGACCGTGTGTCTTCGGACGACCGGGGTGTGTGCGCGGACGGCCAC 60
OY 61 atgaagtttcaagctcatccagcagcggtgtgacccgtaccggaaagcgacgcaag 120
DB 61 ATGAAGTTTCAAGCTCATCCAGCAGCGGGTGTGACCCGTACCGGAAGCGCATCGCCAA 120
OY 121 gatccaactactgatacagtgatcgtcgttgaaacatgaaatgaaatgaaatgaaat 180
DB 121 GATCCAACACTACTGATACAGTGATCGCTTGGAACATGGAGATGGAGAAATCTAGAC 180
OY 181 cttgatgatacttctgtgtatgtagagacgataaagacagactgtgtacgtgtttgat 240
DB 181 CTTGATGATACCTCTTGTGTATGTAGTAGACAGATAAAGACAGACTGTGTAGAGTGTGAT 240
OY 241 gagcagatcccaatacgaacgaggtgtatgtgacagcggtccaggttccacggtaccag 300
DB 241 GAGCAGATCCCAATACGACGAGGTGATGGCACAGTCCAGTCCACGGGTACCCAGAGC 300
OY 301 ccagagatattgttagttagcttggcaccacaatgtctcaagcttcaagcttaccacaa 360
DB 301 CCAGAGATATTGTGTAGTGTAGTGTAGCTTGGCACCAATGTCTCAGCTTTACCTTACCA 360
OY 361 gaacaagtgaatltgagtgacacactcaagcttcctcgaacaaatgtccttccatgt 420
DB 361 GCACAAGTGAATLTGAGTGACACACTTCACTCTTGACCAATATTCCTTCAATGT 420
OY 421 cgaagcagtagttagcccaagctcctaatgtgcttccactctgtatgtatgtatgtt 480
DB 421 CGAAGCAGTAGTGTAGCCCAAGCTCTTAATGTGGCTTCCACCTTGTGATGTATATTT 480
OY 481 tccctcgaagagcttcaagaaatacccaacgctgtgtcaacaacagctgtcctc 540
DB 481 TCCTCGAAGAGCTTCAAGAAATCCACACGCTGTGTCAACACAGCTGTGCTTC 540
OY 541 aagcagaacactgtcgtggaagtcctaaacccctgcgacagaagaagaatgaanaatcaga 600
DB 541 AAGCAAACTCTCTGGAGTCTTAAACCTCGACAGAAAGTGAANAATCACTACAGA 600
OY 601 agactcccgagatctctgtatgtgttcaacaaattcaagagaagaatgtctgtcgt 660
DB 601 AGCTCTCCGGGATCTCTGTATCTGTCTAACCAATTTCTAGAGACCAATGTCTCTCTG 660
OY 601 agctctcccgagatctctgtatgtgttcaacaaattcaagagaagaatgtctgtcgt 660
DB 601 AGCTCTCCGGGATCTCTGTATCTGTCTAACCAATTTCTAGAGACCAATGTCTCTCTG 660
OY 661 tctctagtgccagtaaccacatgtgtggaagtggtcgtgagaacaagaacagagatga 720
DB 661 TCTCTAGTGCCAGTACCAATGTGTGGCAATGTGCTGAGAAACAAGACAGATGAG 720
OY 721 gatggacagaagaagatacagtcgtgtgaaacctgttgacatgtctgaacaggtttg 780
DB 721 GATGGACAGAAAGAGATTAACATCTGTGTAACCTGTTGACATGCTGACACGGGTTTG 780
OY 781 gagcatatcccaacttctcgtgtatgtatgtatgaagtctgtagaagcccccaagat 840
DB 781 GAGCATATCCCAACTTCTCTGTGATGATGTGTAAGCTCTGTAAGAGTCCCAACGAT 840
OY 841 ggaagggcctctgggaatccatgtagtgccttcaagtgtctggaagcggaacacccctgg 900
DB 841 GGAAGGCTCTGGGAATCCATGTAGTGTAGTGTAGTGTAGAGCGGCAACACCTGGGG 900
OY 901 ttaattagtaaacagatltggaagaaggtgtgtaagctgtgaacatgaaatcctttcgtga 960
DB 901 TTATTAGTAAACGATTTGGGAAGGTGTGTAAGCTGTGAACATGAAATCTTTTCGTGAG 960
OY 961 aatgattgcatgtgtcaggttaatgtagtggtggaaccttcgaataatgaagatttgaa 1020
DB 961 AATGATTGCAATGTGTCAAGATTAATGATGGGACCTTTCGAATGAAGATTTGAACAAGCA 1020
OY 1021 caacatatttcgcaaacgacatgctacacccatatttgcattgtgtcattgtcgtcga 1080
DB 1021 CAACATATTTCGCAACGACATGCTTACACCCATATTTGTTGTCATGTGTCTCTGCA 1080

QY 1081 gcaataaagacagatgatgaacatacccaagtgaagaacaatactatctcaagc 1140
 Db 1081 GCAATAAAGACAGATGATGAACATACCCAAGTGAAGAACAATACTATCTCAAGC 1140
 QY 1141 cgttttagccctgacagccagtaattgcaacaagaagtgtgaacagtgacagcttcac 1200
 Db 1141 CGTTTtagccctgacagccagtaattgcaacaagaagtgtgaacagtgacagcttcac 1200
 QY 1201 acagtgacagacacccagcagcagcagcagcagcagcagcagcagcagcagcagc 1260
 Db 1201 ACAGTGACAGACACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1260
 QY 1261 ctacccatagacgacacccctcggaacacacacacacacacacacacacacacacac 1260
 Db 1261 CTACCCATAGACGACACCCCTCGGAACACACACACACACACACACACACACACAC 1260
 QY 1321 aatgattatgacgactgtgaagcagtgatgaacacacacacacacacacacacacac 1320
 Db 1321 AATGATTATGACGACTGTGAAGCAGTGTGAACACACACACACACACACACACACAC 1320
 QY 1381 aatccagccttaagaaagtlacagaagttgggattcagcactcagcactcagcactc 1380
 Db 1381 AATCCAGCCTTAAGAAAGTlacagaagttgggattcagcactcagcactcagcactc 1380
 QY 1441 acaatagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1440
 Db 1441 ACAATAGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
 QY 1501 caagatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1500
 Db 1501 CAAATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1500
 QY 1561 gacgaatcccaagaagaagtgctgctgctgctgctgctgctgctgctgctgctgct 1560
 Db 1561 GACGAATCCCAAGAAGAAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 QY 1621 agccttcgcttcgcttcgcttcgcttcgcttcgcttcgcttcgcttcgcttcgcttc 1620
 Db 1621 AGCCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1620
 QY 1681 agccagatccagatccagatccagatccagatccagatccagatccagatccagatcc 1680
 Db 1681 AGCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCC 1680
 QY 1741 gatgacacagaggaattctgacattgagtcacattgagtcacattgagtcacattgag 1740
 Db 1741 GATGACACAGAGGAATCTGACATTGAGTCACTTGAAGTCCCTTAATGATTGAGTCAAG 1740
 QY 1801 ctgtgtcagtgacagagtgacagagtgacagagtgacagagtgacagagtgacagagtg 1800
 Db 1801 CTGTGTCAGTGACAGAGTGACAGAGTGACAGAGTGACAGAGTGACAGAGTGACAGAGTG 1800
 QY 1861 gtcagatccattatgag 1860
 Db 1861 GTCAGATCCATTATGAG 1860
 QY 1921 caactgaatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1920
 Db 1921 CAACTGAATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1920
 QY 1981 accctaaagagtgatgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1980
 Db 1981 ACCCTAAAGAGTGTGATGACGACGACGACGACGACGACGACGACGACGACGACGAC 1980
 QY 2041 gcaag 2040
 Db 2041 GCAAG 2040
 QY 2101 ctgcccattgaacagcgttgatgatgagagagagagagagagagagagagagagagag 2100
 Db 2101 CTGCCCATGAAACAGCGTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
 QY 2161 gggattgagggcttgatgatgagagagagagagagagagagagagagagagagagag 2160
 Db 2161 GGGATTGAGGGCTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160

Db 2161 GGGATTGAGGGCTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 QY 2221 gacacagtaaatatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2220
 Db 2221 GACACAGTAATAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2220
 QY 2281 atagaagaatgacagagtgacagtgacagtgacagtgacagtgacagtgacagtgacag 2280
 Db 2281 ATAGAAGAATGACAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 2280
 QY 2341 tcccaatgacagtgacagtgacagtgacagtgacagtgacagtgacagtgacagtgacag 2340
 Db 2341 TCCCAATGACAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAG 2340
 QY 2401 agtgaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2400
 Db 2401 AGTGAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2400
 QY 2461 gaagagattgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2460
 Db 2461 GAAGAGATTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2460
 QY 2521 caattgagattcgttaaaacacagaaatacaaaagacatgagttagttatagctgacag 2520
 Db 2521 CAATTGAGATTCTGTTAAACACAGAAATACAAAAGACATGAGTTAGTTATAGCTGACAG 2520
 QY 2581 actaaactacatcagtgatgacacagaagaagtgcttcacagcagagatgagtgct 2580
 Db 2581 ACTAAACTACATCAGTGTGATGACACAGAAGAAGTGCTTCACAGCAGAGATGAGTGTCT 2580
 QY 2641 tccctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2640
 Db 2641 TCCCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2640
 QY 2701 actttgagtgagtgatgacagcagcagcagcagcagcagcagcagcagcagcagcagc 2700
 Db 2701 ACTTTGAGTGAGTGTGATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2700
 QY 2761 tgcaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2760
 Db 2761 TGCAATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2760
 QY 2821 gatgatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2820
 Db 2821 GATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2820
 QY 2881 gacgtcgtatccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2880
 Db 2881 GACGTCGTATCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2880
 QY 2941 aaccaaagaaagtgatgagcagcagcagcagcagcagcagcagcagcagcagcagcag 2940
 Db 2941 AACCAAAGAAAGTGTGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2940
 QY 3001 aaagatgagatgag 3000
 Db 3001 AAAGATGAGATGAG 3000
 QY 3061 gggagacagtgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3060
 Db 3061 GGGAGACAGTGTGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3060
 QY 3121 ataaataacaggaatccttgaacacagaagaagagagagagagagagagagagagagag 3120
 Db 3121 ATAAATAACAGGAATCTTGAACACAGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
 QY 3181 gagaagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3180
 Db 3181 GAGAAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3180
 QY 3241 gctgaatcaggaatccttgaacacagaagaagagagagagagagagagagagagagagag 3240
 Db 3241 GCTGAATCAGGAATCTTGAACACAGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240

Db	3241	GCtAAATTCaAGATTTTCATCCGAGCAATTGGCTGTGATGAGTATATGATATGCGGGGA	3300
Qy	3301	gtctctctctatgaaggttccatggtctctcaagcttagacctcagaagccacagaaggg	3360
Db	3301	GTTCCTCTCTATGAAGGTTCATGGCTCTCAACGCTTAACCTCAGAGCCCAAGAAAGG	3360
Qy	3361	catgtatgagtgtctgttataggcccaagttaagaagccgggaattccaaacccctcact	3420
Db	3361	CATATGATGAGATGCTTTGTATGCCCCAAATCAAGAACCGCGGGAATTCGAAACCCCTCACCCT	3420
Qy	3421	gtagaacgttaacagatcaactctcttagcaatcatgtatcgatatacaagcgtctaggaagaa	3480
Db	3421	GTAGACATTAACAGATCAACTCTTAGCAATCATGATTCGGATATACACCTCTCAGGCAATAA	3480
Qy	3481	ttcaagcaagaacagaacaaatgaagaatgtagaagaatcgctcgycgacctatagttttag	3540
Db	3481	TTTCAGCAAGCAAAAGCAAGATGAAGATGTAAAGATGCTCGGCGGAGCTATAGTTTGG	3540
Qy	3541	caacccttggccgaacagcagccggtcgagccgagagcggtggtgacactggttctgtgag	3600
Db	3541	CAACCTTGCCGGAACGCACGCGCCGCGAGCAGACCGGGGCACTCGGTGCTCGTGAG	3600
Qy	3601	gtgcagatgacgacgacagcagcagagagagcgcgagagctcccaagcagaagcgccag	3660
Db	3601	GTGCAGATGACGCGGACCGGACGAGAGAGCGCGAGACTCCACAGGCCAGCGCCAG	3660
Qy	3661	tacagctctctgtccttcgcaaaagcagaaataatgcacagctcgtctccacagactcttg	3720
Db	3661	TACAGCTCTGTGCTCGGCAAAAGCAAGAAATGCGACACTGGTCTCCAGAGACTCTTG	3720
Qy	3721	gagcagaactctcctccctgtgggaaggtctccagagttgccaagagaaagcccaagtattcc	3780
Db	3721	GAGCAGAACTCTCCCTCCGCGGGAAGGCTTCCAGATGCGCAAAAGAAACCCAGGTATCCC	3780
Qy	3781	agctaaccaaggtctccagaaagcgctactctggagagacatgtgtcttaacgccaaggtcatg	3840
Db	3781	AGCTAACCAAGGCTCCAGSAAAGGCTACTTGGAAGACATGTGCTTAAAGCCAGGGTCAATG	3840
Qy	3841	ctggaaacctcaggaagctccttcgcgccaagaacagaagcggaagagcaagatgaagaag	3900
Db	3841	CTGGAAACTCAGSAGTCTCTTCGCCAGSAAACAGAGCGGGAAGACAGATGAAGAAG	3900
Qy	3901	cagcctctctcagaggggccagaacatagtactgataaagaagttccagaagcccaagt	3960
Db	3901	CAGCTCTCTCCGAGGGGCCAGCAACTATGACTGTATGAAGAAAGTCCAGAGCCCAAGT	3960
Qy	3961	tagcgcctctcccaaggggccccttcgggaagaatgtgccccctctcttaagtttcg	4020
Db	3961	TAGCGCTCTCCCAAGGGGCCCTTCGCGGAAGATGTGCCCCCTCCCTTCAGGTTTCGG	4020
Qy	4021	agcgtgaacacactcagactccttagaagaaggaagccctctacttctga	4071
Db	4021	AGGCTGAACACACTTCAGACTCTGAGAAAGGAGGCCCCCTCTATTTCTGA	4071
RESULT 3			
AF467003			
LOCUS	AF467003	4062 bp	mRNA linear PRI 07-FEB-2002
DEFINITION	Homo sapiens pattingtoning-defective 3 protein splice variant b		
ACCESSION	AF467003		
VERSION	AF467003.1	GI:18568345	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 4062)		
JOURNAL	Gao, L., Macara, I.G. and Joberthy, G.		
REFERENCE	Multiple splice variants of PAR3 and of a novel related gene, PAR3L, produce functionally different proteins		
	unpublished		
	2 (bases 1 to 4062)		

AUTHORS	Gao, L., Macara, I. G. and Joberty, G.
TITLE	Direct Substitution
JOURNAL	Submitted (10-JAN-2002) Center for Cell Signaling, University of Virginia, PO Box 800577, Charlottesville, VA 22908, USA
FEATURES	
source	1. .4062 location/Qualifiers
gene	/organism="Homo sapiens" /db_xref="taxon:9606" 1. .4062 /gene="PARD3" /note="PAR3"
CDS	1. .4062 /gene="PARD3" /note="PAR3: multi-PDZ domain containing protein; alternatively spliced" /codon_start=1 /product="partitioning-defective 3 protein splice variant b" /protein_id="AAU76043.1" /db_xref="GI:18568346" /translation="MKTYVCGRTRVVVPVCGDGHMKFSLIQAVTRYRKAIAKDPNVT WIVYRLRHEGDGGLLDLDLIDCDVADKDRLVAVEEDDPHHGDGASSTGOSPE IFGSELSTNYSARQYQATSEIETVSYLANMLYHRSDDPALIGTSVSDNMR SSEPSRKNTPTRMSTTGAFLKONTAGSEPTCKDRKDEYRSLPDTSMWSQFORDNK RSSLTASHPNWGVKMLEKOEDEDGTEEDNSVVEVGHADTGLEHLPNLSDDVKLVE VPNQDPLGIHVPEPSARGFTLGLRLKKGKAHEHNLFRENDCIVRLINDGLVLR RFEQAOAHFMRQAMRTPIIMEHVPLANKEQEGLDSEKNYSSEPSQYIDNRN SVNAGLHTVQARAPRLNHPPEQIDSHSLPHSAHPSGKPPAPASAPQNVSTTVSSG YNTKIKRRLNIOLKKTGEGFSITSRDYITGSAPLYKNILPRCAIIDGRILKAG DLRLVGVNDLVGKSQSEVVSLLKSTMEGVSLVFRQEDAFPREINAPESOMQIT KETRAEDEDIVLPDGTGREPLTFEPVPLNDSSAGLVKSNRKNENHADGIPVKS INGGAASDGLRLVNPDLIANGESLIGLQMDAMELRLRMSTEGNKRGMILQILVAR RISGCELNKSPSGPPELPIETALDRERISHSIXGIEGLDESPRNALSRIMG KYOLSPYLMSPDQDTVIIEEDRLVPLPHSLDSSSSHDVGYTADGWAARATIS DSACSLSPQDVPLAFORREGRGQSKSEKTKQFSPASQDLPYKTKRSKMDGIAD EKTAKTVDQKAGSPSRDGLKTKSSLSLEQIAAEVTLGDIPEHPRRLIR GRGTNESFRAIDKSYDKPAVDDEDEMETLEEDTESSRGRSVASDQPSHSLTE RONMGNEGKDTDRKDKTKGKEKKRKDKDKMKAKGMLKGLDMFRGKRRKDD KIEATGKIKIOESTSEERIRMKOEGERIOAKREFPERARREDAEIODFRTGEG CDDLEMTGVSYSYGSAALNRPQSPREGHMDPLAIOVKPRNSKSPDQNSSTPS NHDRIOLRQEFQAKODEVEDERRTYSFEQPPNPARPAQOSRHSYSVYQAKQOR QEEBESSQAKQRTYSLPRQSRKNASSVSQDSWQCNISPGGFIQSAKSNPYSSTYG KNGVIGDFENARVMELOELRLQROARKEDQMKOPPEBSPSNYDSKRVQDPSYAR PKGPRQDVPEPSPQVARNLRILQTPKEGRPEYS"
misc_feature	814. .1074 /gene="PARD3" /note="Region: PDZ domain"
misc_feature	1378. .1638 /gene="PARD3" /note="Region: PDZ domain"
misc_feature	1765. .2028 /gene="PARD3" /note="Region: PDZ domain"
BASE COUNT	1218 a 922 c 1075 g 847 t
ORIGIN	
Query Match	71.0%: Score 4041.4; DB 9; Length 4062;
Best Local Similarity	99.8%: Pred. No. 0;
Matches 4061; Conservative	0; Mismatches 1; Indels 9; Gaps 1;
1	atgaagaatgacgcgtgtgttcggaaggaacccgggtgctgtcgttcgctgsggggaagcgaac 60
Db	1 ATGAAAGAGACCCGCTGCTTGAGACGGACCGGGGTGCTGTCCTGCTGGGGAGCCGCCAC 60
QY	61 atgaagaatgacgcgtgttcggaaggaacccgggtgctgtcgttcgctgsggggaagcgaac 120
Db	61 ATGAAAGAGACCCGCTGCTTGAGACGGACCGGGGTGCTGTCCTGCTGGGGAGCCGCCAC 120
OY	121 gatccaaactctcgtgatacagtgatcgtcttgaaacttgaaagtgaagatacgaatactagac 180
Db	121 GATCCAAACTCTCTGATACAGGTGATGCTTGGAAACTGAGATGAGGAAATACCTAGAC 180

QY	181	ttttagatgacattctttgtgatatgacagatataagaacagactggtgtagcagtggttat	240
Db	181	cttgatgtacattctttgtgatgtagtgcagacagatataagaacagactggtgtagcagtggttat	240
QY	241	gagcaggatccacatacacaacgaaggtgtagtgcagacagtgccaggttccacaggtaccacagag	300
Db	241	gagcaggatccacatacacaacgaaggtgtagtgcagacagtgccaggttccacaggtaccacagag	300
QY	301	ccagaagatattgtagttagacttgcagacacaatgtctccagctctttaagctcttccaa	360
Db	301	ccagaagatattgtagttagacttgcagacacaatgtctccagctctttaagctcttccaa	360
QY	361	gcaacaagttaaattaggtcacacatttcagctcttcgcagcaaatatgtcctcttcagt	420
Db	361	gcaacaagttaaattaggtcacacatttcagctcttcgcagcaaatatgtcctcttcagt	420
QY	421	cgagcagtagttagaccagcttaatttggccttcacactctgttagttagttagtaattt	480
Db	421	cgagcagtagttagaccagcttaatttggccttcacactctgttagttagttagtaattt	480
QY	481	tccctcgaagagagccttcagaagaaaaatccacaagctgtgtcacaacagcttgcctctc	540
Db	481	tccctcgaagagagccttcagaagaaaaatccacaagctgtgtcacaacagcttgcctctc	540
QY	541	aagcagaacactgtgtggaggtcctaanaacctgcagacagaagaagatgaaaacctacaga	600
Db	541	aagcagaacactgtgtggaggtcctaanaacctgcagacagaagaagatgaaaacctacaga	600
QY	601	agcctcccgccggagatacagtagtgcgtctaaccaattccagaagacaaatgctgcgtcg	660
Db	601	agcctcccgccggagatacagtagtgcgtctaaccaattccagaagacaaatgctgcgtcg	660
QY	661	tctctgagtgccagtcaccccaattggttggcgaagtgtgcgtgagaaacagaacagatagag	720
Db	661	tctctgagtgccagtcaccccaattggttggcgaagtgtgcgtgagaaacagaacagatagag	720
QY	721	gatggagacagaagaggaataacagtcggtgttgaaacctgttgacaatgctgcagacggtgtg	780
Db	721	gatggagacagaagaggaataacagtcggtgttgaaacctgttgacaatgctgcagacggtgtg	780
QY	781	gagcatataccaaactttctcttgcagatatagttlaaagctgcgtlaagaatgcccaacgat	840
Db	781	gagcatataccaaactttctcttgcagatatagttlaaagctgcgtlaagaatgcccaacgat	840
QY	841	ggaagggccttcgggaatccatgtagtgcttcaagtctcgaagcggcagaacctcggg	900
Db	841	ggaagggccttcgggaatccatgtagtgcttcaagtctcgaagcggcagaacctcggg	900
QY	901	ttattagtaaacagatttggaaaggttgcgtlaaagctgcgaacatgaanaacttttgcgtgag	960
Db	901	ttattagtaaacagatttggaaaggttgcgtlaaagctgcgaacatgaanaacttttgcgtgag	960
QY	961	aatgattgacatttgaagattatgtagtgcagacttcgaatatagaagatttgaacaaga	1020
Db	961	aatgattgacatttgaagattatgtagtgcagacttcgaatatagaagatttgaacaaga	1020
QY	1021	caacatagtttcgcgaagcagatgcgtlaaccaatcaatttgcgttgcgtgcgtgcgtgcgtgc	1080
Db	1021	caacatagtttcgcgaagcagatgcgtlaaccaatcaatttgcgttgcgtgcgtgcgtgcgtgc	1080
QY	1081	gcaaatgaagagcatatgaanaactatccaaagttagaagaagaacattactattcaagc	1140
Db	1081	gcaaatgaagagcatatgaanaactatccaaagttagaagaagaacattactattcaagc	1140
QY	1141	cgatttagccctgcagcagcagatatactgacaacaggaatgtagaacagtgcagggcttcaac	1200
Db	1141	cgatttagccctgcagcagcagatatactgacaacaggaatgtagaacagtgcagggcttcaac	1200
QY	1201	acgggtgcagagagacaccccgactgaaacaccccgctgcagagatagagcttccactcaaga	1260
Db	1201	acgggtgcagagagacaccccgactgaaacaccccgctgcagagatagagcttccactcaaga	1260
QY	1261	ctacctcaatagcacaacccctcgggaaaaaccacatccgcgtccagcctgcgaacctcag	1320

Db	1261	CTACCTCATAGCCACACCCCTGGGAAAACACCATTCCGTCACAGCTCGGACCTCAG	13
QY	1321	aatgatattagtaagacttlaagcagtggttatcaacccaaaaaataagcaagagctt	13
Db	1321	AATGATTTTAGTACGACTGTAGCGACTGCTTATTAACACCAAAAAAATAGCAAGAGCTT	13
QY	1381	aaatcaccaacttaagaagaagtaacgaaggtttggtatcagcatcactccagatgta	14
Db	1381	AATATCACACTTATMAAAGGATACAGAAAGGTTGGATTGAGATTCAGATCACTTCAGAGATGTA	14
QY	1441	acaataagtygcttcagcccaatcatgatgttaaaaaactttccccggggggcgccact	15
Db	1441	ACAATATGGTGGCTCAGCTCCATCTATGTGAAAACATTTCTCCCGGGGGCGGCCATT	15
QY	1501	caggtatgcgaacttaagcagcagcagacagactatagaaatgaatgagtagattlaag	15
Db	1501	CAGGATGGCCCACTTAAGCGAGCAGACAGACTTATGAGGTAAATGAGTACATTTTAGTG	15
QY	1561	ggcaatccccaaggaggaagtggtttgctgtgttagaagccaagaatgaaagaaactgtg	16
Db	1561	GGCAATCCCAAGAGGAAGTGTGTTTCCGCTGTGTAAGAGCACCAATATGGAAGAACTGTG	16
QY	1621	agccttcgtctcttcgcagcagaagacgcttcaccaccaaggaaacttaatgcagagca	16
Db	1621	AGCCTTCGTGCTCTTCGCCAGGAAGACGCTTCACCCAAAGGAACCTAATGCGAAGCA	16
QY	1681	agcagaatgcagaattccaagaagaagaagaacgaagaatgaagatatgtcttacaact	17
Db	1681	AGCCAGATGCAGATTCCTCAAAAGAAAGAAAGAGAGAGATGAGATATGTTCTTAACACT	17
QY	1741	gatggcaccaggaattctcgaatttgaagtcaccaacttaagttatcagtgatgcgcgcg	18
Db	1741	GATGGCACCGGAATTTCTGCATTTGAAGTCCACTTAATGATCAGATCTGCGCAGGC	18
QY	1801	cttggtctcagtgctcaagaaggttaacggtcaaaagaagaacacgcagaaatttggagacttt	18
Db	1801	CTTGGTCTCAGTGTCAAAGGTAAACCGGTCAAAAGAAACCCAGCAGATTGGGAATCTTT	18
QY	1861	gtcaagtcgaattatgaagaagcagacatcaagaatgaagagcttcgggtgaaatgat	19
Db	1861	GTCAAATGCATTATTAATGAGAGACACACTTAAGATGGAAGGCTTCGGGTAATGAT	19
QY	1921	caactatagcagtaatgaggaatccctgttggcagaagcaaaccaagatgcatggyaa	19
Db	1921	CACCTATAGCGATTAATGAGAAATCCCTGTGGCGAAAGCAAAACCAATGTCATGGA	19
QY	1981	acccttaagaagttctatgtctcactgaagcgaataaagaagaaatgatacgaactatgtt	20
Db	1981	ACCCATAAGAGGTCTATGCTACAGACGCATTAACGAGGAATGATCCAGCTAATTTGTT	20
QY	2041	gcaagaagaataaagaagaatgcaatgaagctgaagtcacacttggagcccccctgaaactgag	21
Db	2041	GCAAGAGAAATTAAGCAATGTGCAATGAACCTGAAGTCACTGGAGCCCCCTGGACTGAG	21
QY	2101	ctggccatttgaagaacgcttgatgatgatagagaagcaagatttcccatctccctacagt	21
Db	2101	CTGCCATTGAACAGGCTTGATGATGATAGAAAGAAAGAAATTTCCATTCCCTCTACAGT	21
QY	2161	gggatttggggcttgatgaatcgcccagcagaatgctgcgcctcagtagaataatgtgt	22
Db	2161	GGGATTGGGGCTTGATGAATCCCCAGCAGAAATGCTGCCCTCAGTAGGATAATG---	22
QY	2221	gagtcaggttaataacacagcgtgtccctcaagaggaatatgccccaaagatgaacgtgcaat	22
Db	2218	-----GGTAAATACACAGCTGTCCCTTACAGTGAATATGCCCCCAAGATGACACGTGATT	22
QY	2281	atagaagaatgaaggttgcacagtgtcttcacacatctctgcacagatctctccagc	23
Db	2272	ATAAAGATGACAGGTGTGCAGAGTCTCTCCACATCTCTGTGACAGTCTCTTCCAGC	23
QY	2341	tcacataataatgaggttttgaagcgcagatcgtgtacttggcacaagctgcgaac	24
Db	2341	TCACATAATAATGAGGTTTTGAAGCGCAGATCCTGTACTTGGCACAAGCTGCGAAC	24

Db	2332	TCGCATGATGATGGGGGTTTGGACGGCAGATGCTGGTACTTGGGGCCAAAGGCTGCATC	2391
Oy	2401	agtgatccagccgcagctgcctctttagtccagatgltatccagttcttgcttcttcaagca	2460
Db	2392	AGTATTTACAGCCGACCTGCTTTGAGTCCGAGATGTGATCCAGTTCTTGCTTTCAACGA	2451
Oy	2461	gaagatttggaagtcagtagtactcaagaataacgcacaagaacatttccagttccagt	2520
Db	2452	GAAGGATTTTGGACCTCGAGATGATGAGAAAAACGCACAAACCAATTTTCAATGCCAGT	2511
Oy	2521	caattgatttcgtctaaacacgaaatacaaaagaatgacttaagttatagctgaagag	2580
Db	2512	CAATTTGATTTTCTTTAAACACCAAAATCAAAAAGCATGGATTTAGTACTACAG	2571
Oy	2581	actaacctcaatacaagttggaatgacagaaagcaggtctccccaagagagatgtygtcct	2640
Db	2572	ACTTAACCTCAATACAGTGGATGACAGAAAGCAGGTTTCCACGACGAGATGTGGTGCT	2631
Oy	2641	tccctgggtctgaaagaatcgaagctcgtttgagagatctgcagaccgagttgcgaagtg	2700
Db	2632	TCCCTGGGTCTGAAAGATGCACAGCTCGTTGGAGAGCTGCACACGCAATTTGCCAGGTG	2693
Oy	2701	actttgaatgggaatctctcttccatctgtccacggccgcgagatatcaagagcagggga	2766
Db	2692	ACTTTGAATGGGGATATTTCTTTCCATCTGTCACGGCCCGGATATATCGACGAGGGGA	2751
Oy	2761	tgcgaatagagcttccagaagctgcacatccgaacattctatgataaaccgcgtagatgat	2820
Db	2752	TGCAATGAGACCTTCAAGCGTGCCTCAGACAAATCTTATGATTAACCCCGSGTAAATGAT	2811
Oy	2821	gatgaaagagcaltgagagaccttggaaagaacacacagaagaagtccaagtaagagaga	2880
Db	2812	GATGATTAAGCGCATGAGGACCTTGGAAGAAAGACACAGAAAGAAAGTTTCAAGTACAGGAGA	2871
Oy	2881	gaagcttatcccaagccagatgataagccctcccatctcttgagagccaatgtgata	2940
Db	2872	GAGCTGTATCCACAGCCAGCTGATCGACCTCCCATCTCTGTGAGAGCAAAATGTAGTGA	2931
Oy	2941	aaccaagagaaagttgtataagactgtatagaataaagaataaaactcgtgaaagaagaag	3000
Db	2932	AACCAAGAAAGTGATTAAGACTGATGATGAAAAAGATTAACCTGCAAAACAAAGAGAG	2991
Oy	3001	aaagatagagataagagaagaagataaatagaaagccagaagaagaaatgctgaagggcttg	3060
Db	2992	AAAGATGAGATTAAGGAGGAAGATTAATAAGTAAAGCCAAAGAGGAATCTGTCAAGGCTTGG	3051
Oy	3061	ggagacatgttcaggtttgtgccaacatcgaaaagatgacaagaattggagaagaacggtataa	3120
Db	3052	GGAGACATGTTCCAGGTTTGGCAAAACATCGAAAAGATGACAAAGATTGAGAAAAACGGGTAAA	3111
Oy	3121	ataaaatacagaagatcccttaccataagaagagagagatacgaataagacagaagag	3180
Db	3112	ATAAAAATACAGGAATCTTTTACATCGAAGAGAGAGATTCGANTGAACGACGACAG	3171
Oy	3181	gagagatccaagccaaaactcgagaatttagyaaacgaacgactcgagaagcgttactat	3240
Db	3172	GAGAGGATTTCAAGCCAAAACTCGAGAAATTTAGGAGACGACAAAGCTCAGAGCGTGACTAT	3231
Oy	3241	gctgaaatcaagatttctatcggaacattggtcgtgtatgtagagtaatgataatgagggga	3300
Db	3232	GCTCAAAATTCANAGTTTTCATCGGACATTTGGCTGTGAAGATGAATGATATGTTGGGGGA	3291
Oy	3301	gttctctctataaggtttccatgagctctccaagctcagaagcccaagcccaagagaag	3360
Db	3292	GTTTCTTCTTATGAAGTTTCATAGGCTCTTCACGCTAGACCTTCAGAGCCCAAGCAAGG	3351
Oy	3361	catatgaltgagttctgtatgtgccaaagtcgaagaagccgcyggaattccaacccctcact	3420
Db	3352	CATATGATGAGTCTGTTGTATGGCCCAAGTCAAGAACCCCGGAATTTCAAAACCTTCACCT	3411
Oy	3421	gtagaacgtatacagaataactcctagaaatcatgatctgatacgaagcttcgaagagaa	3480
Db	3412	GTAACAGTAAACAATCACTCTTAGCAATCAAGATCGATACAGGCTGTGAGGCAAGAA	3471

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	CDS
AF467004	3960 bp mRNA linear	AF467004	1	GI:18568347	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	Gao,L., Macara,I.G. and Joberty,G.	Multiple splice variants of PAR3 and of a novel related gene, PAR3, produce functionally different proteins	Submitted (10-JAN-2002) Center for Cell Signaling, University of Virginia, PO Box 800577, Charlottesville, VA 22908, USA	<p>1. 3960</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>1. 3960</p> <p>/gene="PAR3"</p> <p>/note="PAR3"</p> <p>1. 3960</p> <p>/gene="PAR3"</p>	

FEATURES	Location/Qualifiers
source	1..3960 /organism="Homo sapiens" /db_xref="taxon:9606"
gene	1..3960 /gene="PARD3" /note="PARD3"
CDS	1..3960 /gene="PARD3"


```

OY 1501 caggatggcgaacttaaggcagagacagacttataaggltaaatlygagtagatagtg 1560
DB 1501 CAGGATGGCCGACTTAAGGCGAGGACAGACTTATAGAGTAAATGAGATGATTAGTG 1560
OY 1561 ggcgaatcccaaggaggaagtgttctgctgtgttgagaagacccaagaatgagaagacttg 1620
DB 1561 GGCCAATCCCAAGGAGGAAGTGTGCTGTGTGAGAACACCMAATGGAAGGAAGCTGTG 1620
OY 1621 agccttcgtctcttcgcaggaagacgcttccaccagggaaactgaatgacagagca 1680
DB 1621 AGCCTTCGGTCTTTCGCGCAGGAAGACGCTTCCACCAGGGAACCTAATGACAGGCCA 1680
OY 1681 agccagatgacagatcccaaaagaaacgaaagcagaagaatgagatattgttcttaacct 1740
DB 1681 AGCCAGATGACAGATTCCAAAAGAAACGAAGACAGAAATGAGGATATTGTTCTTACACT 1740
OY 1741 gatggcaccaggaattctctgacatttgaatgccacttaatgatacagatctgcaagc 1800
DB 1741 GATGGCACCGGAATTTCTGACATTTGAGTCCACTTAATGATTCAGGATCTGCAGGC 1800
OY 1801 ctgtgtcgaagtgtcaaaagtaaccggtccaaagaagacccagcagatttggaactctt 1860
DB 1801 CTGTGTGTCAGTGTCAAAAGTAAACCGGTCCAAAGGAACACGCGATTTGGGAATCTTT 1860
OY 1861 gtcaagtcattatattatlygaggaagcagacatctaagaatggaaggcttcggttgaaatg 1920
DB 1861 GTCAAGTCCATTATTAATGAGAGACACACATCTAATAAGATGGAAGGCTTCGGGTGAATGAT 1920
OY 1921 caactgatagaatgaatgaggaatccctgtttggcagaagaacaaaccaagaatgccaatgaa 1980
DB 1921 CAACGTATGAGCATGAATGAGGAATCCCTGTTGGGCAAGACCAAAACCAAGATGCCATGGA 1980
OY 1981 accctaagaaggtctatgtctactgaaggcaataaagcaggaatatacagcttatgtt 2040
DB 1981 ACCCTAAGAAGGTCTATGTCTACTGAAGCAATAAACGAGATATATCAGCTTATTGTT 2040
OY 2041 gcaagagaataaagcaagtgcgaatgcaatgagctgaatgcaactggaagccccctlygacctgag 2100
DB 2041 GCAAGAGAATAAAGCAAGTGCATGCAATGAGCTGAAGTCACTGGGAGGCCCTCGACCTGAG 2100
OY 2101 ctgcccattgaagaacgcttggatgataagagaacgaagaattcccatctcccttaacagt 2160
DB 2101 CTGCCATTGAAGAACGCTTGGATGATAGAGAACGAAGAATTTCCCATTCCTTCACACT 2160
OY 2161 ggaattgaggggcttgatgaaatcgccagcagagaatgtcgtccctcagtagagaataatggt 2220
DB 2161 GGAATTGAGGGGCTTGATGAATCCCCCAGAGAAATGCTGCCCTTAGTAGGTAATGGGT 2220
OY 2221 gactcaggtlaaataacacagctgtcccttaacatgaaatgcccccaagatgacactgtcat 2280
DB 2221 GAGTCAGGTAAATACCACTGTGCCCTACAGTGAATATGCCCAAGATGACACTGTCAAT 2280
OY 2281 atagaagatgacaggttcccgatgtcttccacatactctcgcagacagctccctcccg 2340
DB 2281 ATAGAAGATGACAGGTTCGCGATGCTTCTCCACATCTCTGACACAGTCCCTTCCACG 2340
OY 2341 tcccatgatagtgtggtttgtgacgagcagaatgctgtacttgggccaaggtcgaatc 2400
DB 2341 TCCCATGATGATGTGGGTTGTGTGACGGCAGATGCTGTGACTTGGGCCAAGGCTCGAATC 2400
OY 2401 agtgaatcagcagactgctcttgaatccagaatgttgatccagatctctgtcttcaaga 2460
DB 2401 AGTGAATCAGCGACTGCTCTTGAATCCAGATGTGATCCAGATCCAGTCTTCAACGA 2460
OY 2461 gaaagattgagcgtccagaatgctccagaaaaaacgacaagaatcttcaatgccagt 2520
DB 2461 GAAAGATTGAGCGTCCAGAAATGCTCCAGAAAAACGACAAAGCAATTTCAATGCGCAAT 2520
OY 2521 caattggaatttcgttaaaacagaaatcaaaaacatgtaatttaagtaatactgacag 2580
DB 2521 CAATTGGAATTTGTTAAACAGAAATCAAAAACATGATTTAGTATTAAGTATGACGAG 2580
OY 2581 actaaactcaatlacagtgagatgaacagaagaagcaggttctccacagagatgttggctct 2640
DB 2581 ACTAAACTCAATLACAGTGAATGACAGAAAGCAGGTTCTCCACAGAGATGTGGGTCTCT 2640
OY 2641 tccctggcttgaagaagtcaagctcgtttggaatcttcgagaccgagcttgccaggtg 2700
DB 2641 TCCCTGGCTTGAAGAAGTCAAGCTCGTTGGAGATCTGCAACGCGCATGTTCCAGGTG 2700
OY 2701 acttgaatlyggaatcttccatctcgcagcgcccgagtaatacagaagcagggga 2760
DB 2701 ACTTGAATGGGGAATATCTCTTCCATCTGCTCCACGCGCCGGAATATCAGAGCGCAGGGGA 2760
OY 2761 tgcgaatgaagcttcaagcttgccatcgacaacttatalyalaaccccgyltagatg 2820
DB 2761 TGCATGAGAGCTTCAAGCTTCCATCGCAATCTTATGATTAACCCCGGTAGATGAT 2820
OY 2821 gatgtatgaagcatgagaccttggagaagaagcacagaagaagttcaagaatcaaggaga 2880
DB 2821 GATGATGAAGCATGAGACCTTGGAGAAGACACAGAAAGAAAGTTCAAGATCAGGAGA 2880
OY 2881 gagtctgtacacagccagtgatcagccttccactctcttgagagacaatgaaatgaa 2940
DB 2881 GAGTCTGTATCCACAGCCAGCATGATCAGCTTCCCACTCTCTGGAGAGCAATGAAATGGA 2940
OY 2941 aaccagaagaagtgataagactgtatagaataaagaataaaactygaagaagaagaag 3000
DB 2941 AACCAAGAAGAAAGTGATAGACTGTAGAAAAAAGATAAAACTGCAAAAGCAAAAGAAAG 3000
OY 3001 aagaatagaatgaagaagaagaataaagaagaagaagaagaagaagaagaagaagcttg 3060
DB 3001 AAGATGAGATGAGATGAGAGAGATGAATGAAGGCCAAGAGAGGAATGCTGAAGGGCTTG 3060
OY 3061 ggaagcatgttcaaggtttgcaaacactcgaagaagaatgaaagaatgagaagaacgyltaa 3120
DB 3061 GGAGCATGTTC----- 3072
OY 3121 ataaaaatacaggaatccttatacagaagaagaagaatagaatgaagcagagcag 3180
DB 3073 ----- 3072
OY 3181 gagaagattcaagccaaactctgagaatlttaaggaaagcagcttcgaagcgtgactat 3240
DB 3073 ---AGATTCAAGCCAAACTCGAGAAATTTAGGGAACGCAAGCTCGAAGGCTGACTAT 3129
OY 3241 gctgaatccaagatttcatcgcgaacttggctgtgagatgagatgaatgtatgagggga 3300
DB 3130 GCTGAATCCAAGATTTTCAATGCGACATTTGGCTGTGATGATGAGTAATGATGGGGA 3189
OY 3301 gtctcttctatgaaggttccatggtcctcaacgcttagacccctcaagagcccaagaaagg 3360
DB 3190 GTTCTCTTATGAAGGTTCATGCTGCTCAACGCTAGACCTCACAGCCACGAGAAAGG 3249
OY 3361 catatgatgtgacttctgtatcccaagtcagaagcccggaatctccaaacctact 3420
DB 3250 CATATGATGATGCTTGTGATGCCCCAAGTCAAGAAACCCCGGAATTCAAACCTCTACT 3309
OY 3421 gtagaacgaatacagatccaactcttagcaatcaatgatacagatcaggttgaagcaaga 3480
DB 3310 GTAGACAGTAAACAGATCACTCTAGCAATCAATGATGCGATACAGGCTGTGAGCGAAGA 3369
OY 3481 tttcagaagaagaagaatgaagaatgataagatcgttcggcggaactatagtttgaag 3540
DB 3370 TTTCAAGCAAGCAAAAGCMAATGATGAAGATGCTGCGCGGACCTATAGTTTGAAG 3429
OY 3541 caaccttggcgaacgacgagcgagcaggaagcagagcgcagacactcgttccgtyggag 3600
DB 3430 CAACCTTGCCCAACGACGACGCGGCGAGCGCAAGGCGGCGAAGCTGGTGTCCGTGGAG 3489
OY 3601 gtgcagatgcagcgcgcagcgcagagagagcgcgcgaagatcccaagagcccaagcgcag 3660
DB 3490 GTGCAAGATCAGCGGAGCGGAGGAGGAGGAGCGGAGAGCTCCACAGAGGCCACGCGCGAG 3549
OY 3661 taagctctctcgtcgcagcaagaagaagaatgccaagctcggctcccaagactcttgg 3720
```

[illegible]

```
QY 601 agctcccgcgagatactagtaactggtcttaaccaattcagagagaatgctgcgtcg 660
Db 601 AGCCCTCCGCGGAGTACTAGTAACGTGCTTAACCAATTTTCAGAGAGACAAATCTCGCTCG 660
QY 661 tctctgagtcgacgtcacccaatggtggtgcaagtgtgtgagaaacaagaatgagtg 720
Db 661 TCTCTGAGTGGCGAGTCACCAATGGTGGGCAAGTGGCTGGAACCAACAGAGTGTAG 720
QY 721 gatggagacaagaagatacaacagtcgtgtgaacctgtgtgagacatgctgacacgggttg 780
Db 721 GATGGAGACAGAGAGATATACAGTGTGTGTAACCTGTGGACATGCTACACAGGGTTTG 780
QY 761 gagcatataccaactctctctgagatgatalgtlaaagctctgtagaagtcgcccaagat 840
Db 761 GAGCATATACCAACCACTTCTCTGATGATATGTAAAGCTGTGAAATCTCCCAACGAT 840
QY 841 ggaaggacctctgagatcaatgtagtgccttcagtgcctgcagagggagaaacccctggg 900
Db 841 GGAGGGCCTCTGTGGAAATCCATGTAGTGCCTTCAAGTGTGAGGCGGCAAGACCTGGGG 900
QY 901 ttattagtaaaaacgaatggagaagaagtgtaagctgaaactgaaatcctcttcgtgag 960
Db 901 TTATTAGTAAACGATGTGGAGAAAGTGTAAAGCTGAACGTAACATGTTTGTGTAG 960
QY 961 aatgattgcatgtcagagataatgatagtgcgacctcgaatagaagatttgaacaaga 1020
Db 961 AATGATTTGATTTGATGAGATTAATGATGCGCACTTCGAATATGAAAGATTTAAACAAGCA 1020
QY 1021 caacatagtttcgcaagacgaatgctgataccatcattgtgttcctgtca 1080
Db 1021 CAACTATGTTTCCGCAAGCCATGCGTACACCCATCATTTGTTCCATTTGGTTCTGTGA 1080
QY 1081 gcaaaataaagacagatagaaactatcccaagtgagaagaacatctactatcaacg 1140
Db 1081 GCAAAATAAGACAGATATGAAACACTATCCCAAGTGAAGAACAATTAATTCATTCAAAC 1140
QY 1141 cgtttagcctcgagacagcagatataatgacaagaagatgtgaaacgtgcagagctcag 1200
Db 1141 CGTTTACCCCTGAGCGCACAGTATATGTGACAAAGAGTGTAAACAGTCCACGGCTTCAAC 1200
QY 1201 acggtgcagagagaccccgactgaaaccccgctgagcagatagactctactcaaa 1260
Db 1201 ACGGTGCAGAGAGACCCCGACTGAACACCGCTTGACCAATGACTCTCACTCACTAACA 1260
QY 1261 ctacctctaaagcacaacccctcgagaaacacacatccgctcagcctcgcgcactcag 1320
Db 1261 CTACCTCAATAGCGCACACCCCTCGGAAACACCAATCCGCTTCAGCCCTCGCACCTCAG 1320
QY 1321 aatgataattagtaagacgtgtaagcagtggtataacaaccaaaaaataagcaagaagct 1380
Db 1321 AATGATATTAGTACACTGTAAAGAGTGTATTAACACCAAAAAAATAGGCAAGAGGCTT 1380
QY 1381 aataccagcttaagaagaagtgacagaagtttggatcagatcaactccagagatgta 1440
Db 1381 AATATCCAGCTTAAGAAAGTACAGAAAGTTTGGGATTCAGATCACTTCCAGAGATGTA 1440
QY 1441 acaataagttgctcagctcccaactctatglaaaaaacatctcccccggggggcgccatt 1500
Db 1441 ACAATAGTGGCTCAGCTCACTCAATCTATGGAATAAACATTCCTCCCGGGGGCGGCATTT 1500
QY 1501 cagagatgcgcgaacttaaggcagagagacagactctatagagtaaatgagtagatlaagt 1560
Db 1501 CAGGATGGCGCGACTTAAGGCAGAGACAGACTTATAGAGTAAATGAGATATTAGT 1560
QY 1561 ggcacaatcccaagaggaagttcttcgctgtgtgagaagaccaaagatgaaagaagctgt 1620
Db 1561 GGCAAAATCCCAAGAGAACTTTTCGCTGTGTGAGAGACCAAGATGGAGAAAGAACTGTG 1620
QY 1621 agcctctctgtcttcgcagagaaagacgctctcacccaagggaactgagatgcagagcca 1680
Db 1621 AGCCTTCTGTGCTTTCGCGCAGAGAGACGCGCTTCCACCCCAAGGAGAACT 1680
QY 1681 agccagatgcagatcccaaaagaacgaagacagagatgagagatatgttcttaacact 1740
Db 1681 -----GAAAGCGAAGATGAGATATGTTCTTAACACT 1740
QY 1741 gatggaccaggaatcttgaacatttgaagtcgccacttaatatcaagatctgaagc 1800
Db 1741 GATGGACCAAGGAATTTCTGACATTTGAAAGTCCCACTTAATATTAAGATCTGAGGC 1800
QY 1801 ctgtgtcagtgctcaagaaglaacccggtcacaagaagaacacagagatttggaaatcct 1860
Db 1801 CTGTGTCAAGTGTCAAGGTAACCGGTCAAAAGAAACACGACGACATTTGGAAATCTTT 1860
QY 1861 gtcaagtccaattataatgagaagacagacatctaaagatggaagctctgggtgaaatgat 1920
Db 1861 GTCAAGTCCATTAATATGAGAGACGACATCAAGATGGAAGGCTTCGGGTGAATGAT 1920
QY 1921 caactagacagtaaatgagaatccctgtgtggaaagaacaaaccaaatgtcagagaa 1980
Db 1921 CAACCTATAGCACTAATATGAGAAATCCCTGTGTGGCAAGCAACCAAGATGCAATGAA 1980
QY 1981 accctlaagaaggtctatgtctactagaagcaataaaccagagaatgataccagcttatgt 2040
Db 1981 ACCCTAAGAAGGCTATGTCTACTGAAGGCAATAAACGAGAAATGATTCAGCTTATTTGTT 2040
QY 2041 gcaagagaataaagcaagtgaatgagctgaagctcaactggagagcccccctggagctgag 2100
Db 2041 GCAAGAGAAATAGCAAGTGAAGTGAAGTCACTGAGTCACTGGGAGCCCTCGACCTGAG 2100
QY 2101 ctgcccattgaaaacagcgttggatgataagaagaagaattcccatccctcactcagt 2160
Db 2101 CTGCCATTGAAACAGCGTTGGATGATAGATAGAAAGAAATTTCCATTCCTCTCAAGT 2160
QY 2161 gggattgaggggcttgatgaatcgcccagcaagaatgctgcctcagtagtaaatgggt 2220
Db 2161 GGGATTGAGGGGCTTATGATATCCCCAGAGAAATGCTGCCCTCAGTAGGAATATG--- 2220
QY 2221 gagtcaggtaataaccagctgtcccttaacagtgaaatatagcccaagaatgaactgtcatt 2280
Db 2221 -----GGTAAATACCAAGCTGTCCCTACAGTAAATATGCCCAATGACACTGTCAAT 2280
QY 2281 atagaagaatgaaggtgtgcagtgcttccctcacaactctctgaacagtcctcttcacg 2340
Db 2281 ATAGAAGATGACAGGTGTGCCAGTGTCTCTCCACATCTCTTCAACAGTCTTTCACAG 2340
QY 2341 tcccatgatagtgtggggttltgtaacgagatgctgtaactgtgggccaaggtctcaatc 2400
Db 2341 TCCCATGATGATGCTGGGTTTGTGAGCGGAGATGCTGTACTTGGGCCAAAGCTTGCAATC 2400
QY 2401 aatgataacgcgcagctgcctcttgagatccaagatgtgataccagttctgtcttcaaga 2460
Db 2401 AATGATTAACGCGACGACGTCTTGTGAGTCCAGATGTTGATCCAGTCTTGTCTTCAACGA 2460
QY 2461 gaaagattggaagctcagagatgtctcagaanaaacgacaagaatcttcaagatgcagt 2520
Db 2461 GAAAGATTGGAGCTAG----- 2520
QY 2521 caattggaattcgttlaaaacacgaataatcaaaaaagatgattagatagctgcagag 2580
Db 2521 -----ATAGCTGACGAG 2580
QY 2581 actaaactcaataacagtgatgataccagaagaagcaggtctccccaagaagatgtggctct 2640
Db 2581 ACTAAACTCAATAACAGTGAATGACAGAAAGCAGGTTCTCCAGCAGAAATGTGGGTCTCT 2640
QY 2641 tccctgtgctgaagaagcacaagctgtgtggaagatctcagaacgcagatttgcgcagagt 2700
Db 2641 TCCTGTGGTCTGAAGAGTCAACCTGTGTGAGAGTCTGACGCCCAAGTGTGCGAGGTG 2700
QY 2701 acttgaatgggagataatctcttccatcgtcccaagcgccgagataatccagagagcaaggga 2760
Db 2701 ACTTGAATGGGATATATCTCTTCCATCGTCCACGCGCGGATATATCAGAGGCAAGGGA 2760
QY 2761 tgcataagagcttcagagctgcacatcgacaacattatgataaacccgcggtagatgat 2820
Db 2761 ----- 2820
```

Db	2623	TGCATAGAGCTTCAGAGCTGCATGCATCAAAATCTTATGATAAACCCGGGTAGATGAT	2682
OY	2821	gatatgaagcattgagagccttggaaagaacacagaagaagatcagaagagaga	2880
Db	2883	GATGATGAAAGGATGAGAGCTTGGAAAGAACACACAGAAATCTTCAAGATCAGGACA	2742
OY	2881	gagctgtatccacagcagcagatgatalcagcctccacactcttggagagacaaatgatatga	2940
Db	2743	GAGTCTGATCCACAGCCAGCATGATCAGCTTCCACTCTCTGTGAGAGCAAAATGAAATGAA	2802
OY	2941	aacaaagaagaaggatgataagactgataagaaaagaataaaactgaaagaagaagaag	3000
Db	2803	AACCAAGGAAAGGTGATTAAGCTGATAGAAAAAGATTAATACTGAAAAAGAAAGAG	2862
OY	3001	aaagatagagataagagagagagataaatgaaagcaagaagaagaatctcgaaggcttg	3060
Db	2863	AAAGATGAGATAGAGAGAGAGATTAATAAGAAAGCAAGAAAGGAAATCTGAAAGCGCTTG	2922
OY	3061	ggagaacttccagcttctggcaaacatcgaaaagatgagaaagattgagaaaacgggataa	3120
Db	2923	GGAGACATGCTCAGGTTTGGCAAAACATGCAAAATGACAAAGATGAGAAAAACGGGTAA	2982
OY	3121	ataaaatacagaagatccttaccatcgaaagagagagagatacgaatgagagagagcag	3180
Db	2983	ATTAATAATACAGAGATCTTTACATCAAGAAAGAGAGATACGATGAAAGCAGAGAGAG	3042
OY	3181	gagagattcaagccaaactcgagaattagaggacgaagcctcgagagcgtgactat	3240
Db	3043	GAGAGATTCAGCCAAACCTCAGAAATTTAGGGAACGAAACCTCGAGACCGTGCATAT	3102
OY	3241	gctaaattcaagatttccatcgacatttgcgtgatgatgatatgatatgagagagag	3300
Db	3103	GCTAAATTCAGATTTTATCGACATTTGGCTGTGATGATGATTAATGATGAGGGA	3162
OY	3301	gttctcttatgaagttccatgctctcacaagctagaccccgagccagagagagag	3360
Db	3163	GTTTCTCTTATGAAGTTCCATGCTCTGCAACCTGAGACCTGAGCCCAAGAGAGG	3222
OY	3361	catatgatgatgttctgtatgcccagaatcagaagccgagagatcccaacctcaact	3420
Db	3223	CATATGATGATGCTTCTGTATGCCCCAAGTCAAGACCGCGGAAATTCAGAACTTCACCT	3282
OY	3421	gtagaagataaagatacactcactcactcactcactcactcactcactcactcactcact	3480
Db	3283	GTAGACATGATCAAGATCACTCCAGCAATCATGATCGATACACGCTGTGAGCAAGAA	3342
OY	3481	ttcaagaagaagaagaatgagaatgtagaagatcgtcggcggagactatagtttggag	3540
Db	3343	TTTACAGCAAGCAAGCAAGATGAAATGTAAGATCTCGCGGAGCACTATAGTTTGG	3402
OY	3541	caaccctggcgaagcgaagcggcggcggcggcggcggcggcggcggcggcggcggcggc	3600
Db	3403	CAACCCGCGCCGAAAGCAGCGCCCGCGAGCAGAGCGCGGCGACACTGTGCTCGGAG	3462
OY	3601	gttcaagatgagcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc	3660
Db	3463	GTGCAATGTCACACCGCAGCGGAGAGAGCGGACGACAGCTCCACAGCAGCCACGCGCAG	3522
OY	3661	tacagctctgctcggcgaagaagaagaatgacagctcgtctcccaagactcttgg	3720
Db	3523	TACACCTCTGCTCGCGCAAGAGGAAATAATGCAAGCTCGGTCTCCAGAGACTTGG	3582
OY	3721	gagcagactctcctcctggggaagcttccagatgtagcaaaagaagaaccccaagttactcc	3780
Db	3583	GAGGAGAACTACTCCCTCGGGAAGGCTTCCAGAGTCCCAAGAAAGAAACCCAGTACTCC	3642
OY	3781	agctcaagaagctccagaagcgttactcgggagagacattgctcaacggcgaaggtcatg	3840
Db	3643	AGCTACCAAGGCTCCAGAAACGCTTACTGGAGAGACATGCTTCAACGCCCAAGGCTCAG	3702
OY	3841	ctggaatacagaagctctctcgcagaacagaagcgagagagagagagagagagagagag	3900
Db	3703	CTGGAATCTCAGAGAGCTCTTCCAGGAACAGAGCGGGAAGAGAGAGAGATGAAAGAG	3762

OY 3901 cagctctcttcgaggggcccagaactatgactcgtatagaagaatcagaagaccagt 3960
 Db 3763 CAGCCCTCTCCGAGGGGCCAGCAACTATGACTCTCTATTAAGATCAGAGCCAGT 3822
 OY 3961 taccgctctcccaaggagccctccggcaagatgagccctccctcctcctcaggttgcg 4020
 Db 3823 TAGCGCTCTCCCAAGGGGCGCTTCGCGCAAGATGAGCCCTCCCTCTCTCAGATTGCG 3882
 OY 4021 aggttgacagagactcagctcctcgtgaaagaaggagccctctactccta 4071
 Db 3883 AGGCTGAACAGACTTCAGACTCTCTGAGAAAGGAGCGCCCTCTATCTCTGA 3933

RESULT 6
 AF332593
 LOCUS
 DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long
 ACCESSION AF332593
 VERSION AF332593
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCES
 AUTHORS
 TITLE

1 (bases 1 to 3822)
 .Down-regulated expression of atypical PKC-binding domain deleted
 Cell Res. 11 (3), 223-229 (2001)
 11642408
 2 (bases 1 to 3822)
 Fang, C. and Xu, Y.
 Exon/Intron Structure And Splicing Variants of A Novel Human
 Polarity Gene, HASTP
 Unpublished
 3 (bases 1 to 3822)
 Fang, C. and Xu, Y.
 Direct Submission
 Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
 Oncology, Institute of Biochemistry and Cell Biology, Shanghai
 Institutes for Biological Sciences, Chinese Academy of Sciences,
 320 Yue-Yang Road, Shanghai 200031, P. R. China
 Location/Qualifiers
 1..3822

CDs

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /map="10p11.2"
 1..3822
 /note="ASTPb: alternatively spliced; similar to Genbank
 accession number AF196185"
 /codon_start=1
 /product="atypical PKC isotype-specific interacting
 protein long variant b"
 /protein_id="AAK69193.1"
 /db_xref="GI:14579311"
 /translation="MKTVTCGRTRVYVPCGDHMKVFLIQAVTRYKATAPNY
 MIVHRLERHGDGILDDLDIDYADKDLVAVFEDDPHGGGATSSSSQSP
 ITGSELGNVNSAFQYQATSEIVTSVLRAMPPLHVRSSDPALILSTSVSDSP
 NSLSASHPNPTKMTSTAGTAKONTAGSPKCDKRDENRSLPRTSMNSMFOQDVA
 VYNDGPGIIVHPVPSAKGRITGLVLRLEGGGAHEHNTFREDDICVRIINDDLN
 RYNSAGHTYQKAPRLPIWTHVVPANKEOYEOISQEKNNYSRSRSPDSQYIDR
 VNTKTRKRLNIDLKGTGELGFSITSRDVTGSGAPLYVNIIPRGAALIQDSTVSG
 DRLLENGVDLVGKSOEYVGLRSTKMGVSLVLFQEDAPFRLKAEDDYLIVP
 PDGARELVEPLVNDGNSAGIGVYKRSKSNHMDIGIPVKSIIINGAAIDSGDGLR
 VNDLIVAVGESLIGTNDAMDPTLRMSKSTGSKRGMQIOLIVARRISKNSKSGS
 PGPPELPIETALDEGRSHSLTISGIGTESSRNALSRIMGYQSLSPYNDP
 DTVIDEDRLPVLPHLSQSSSSSHDVGFTVADACTWAKAALISADCSLSPVDV

VLAFORSGPFGIIOADETKLMLVDDOKKAGSSRVDGSLGLTKSSLSLSLITATAVEYTL
 NGDLPFHRPFRJIIIRGRCNCSPPAALIDKXVDRKADDDGDEGATLLEPHIDEESSRGR
 ESYVTAADSDPSHSLEROMNGNQEGKDDTDTRKKDXTGKEKKDKRDEKDKAKKAKGMLK
 GLGEMFIQIACTREFREKQAREYDAETLDHRTFGDCDEDEMTGSGVSYESKALMLNR
 POSREGHMLDALTAYQVKPRPNKSPEVDSNRSTPSNHDRIQRLRDEFOAKODEVE
 DRRATYSFDEWPNARPATQSGRHSVSEVQMOQROBERSSQAOARVSLRPELORR
 KNASSVSODSDEWPNARPATQSGRHSVSEVQMOQROBERSSQAOARVSLRPELORR
 ROEORRKEQOQKQPSBEGSPSNYSYDKXVDDPSAPKPGPRDQVPPSPSOVALNMLD
 QTPERKGRPFYS"

1114 a 885 c 1019 g 804 t
 BASE COUNT
 BRIG IN

Query Match	62.0%	Score 3529.8	DB 9	Length 3822
Best Local Similarity	93.8%	Pred. No. 0		
Matches 3820; Conservative	0	Mismatches 2	Indels 249	Gaps 4

OY	1	atgaagatgacccgtgtgcttcgacgcagcccggtgtgtctgtccgttcggggagcgccac	60
Db	1	ATGAAGATGACCCGTGTGCTTCGACGAGCACCGGGGGTGTGTCGCCCTCGGGGACCGCCAC	60
OY	61	atgaagatlttcagcgtcatccagcagccgttgacccgctacccggaaagcgcatcgccaag	120
Db	61	ATGAAGATTTTCAGCCTCATCCAGCAGCCGGTGACCCCTTACCGAAGGCCATCGCCAAG	120
OY	121	gatccaactacttggatacaggatgcatcgtcttggaaactggagatggagaaatactagac	180
Db	121	GATCCAACACTGTGATACAGGTGATCGTCTTGGAAACTGGAGATGGAGAACTACTAGAC	180
OY	181	cttgtatgcattctcttgtgtgtagcagaagataaagacagactgtgtgcagtgtttgat	240
Db	181	CTTGATGCAATTCCTTGTGTGATGTACCAACACATAAAGACAGCTGTGTGACAGTGTATTAT	240
OY	241	gagcagagatccacatcacaggaaggtatgtgacacagtgccagttccacaggtatcccaagac	300
Db	241	GAGCAGATTCACATCAGCAGAGTGATGTGGACACAGTGGCAGTTCCACGGGTATCCCAAGC	300
OY	301	ccaagagatattgtgtagtgagcttggccaaccaaatgtctcagcccttcagccttaccaa	360
Db	301	CCAGAGATTTTGTGTAGTGAGAGCTTGGCACCAAAATGTCTCAGCCTTCAGCCTTACCAA	360
OY	361	gcaacaagtgtgaatgagtgtaacacacctcagtcctctgagaanaataatcctctcatgtt	420
Db	361	GCAACAAGTGAATGTGAAGTGCACACTTCCTCAGTCTTCAGCAAAATATCCCTTCATGTT	420
OY	421	cgaagcagtagtgacccagcctctaaatttggcctctccactctcgttaagtgataaattt	480
Db	421	CGAGCAGTAGTGACCCAGCCTCTAATTGGCCCTCCACTTCGTGAAGTATGTAATTTT	480
OY	481	tcctctgaaagaccttcaagaaaaatcccacacgctgtgttcaacaacagcttgcttcctc	540
Db	481	TCCCTGGAAGACCTTCAAGGAAAAATCCCAACACTGGTCMAACAGCTGGCTCTCCTC	540
OY	541	aagcagaacacactgctgggagttccttaaaccttcgacaggaagaagaatgaaaacctacga	600
Db	541	AAGCAACACACTGCTGGGAGTCTTAAACCTTCGACAGGAAAGATGAAACTTACAGA	600
OY	601	agctcccgcggaatactagtaacttgcgtctaaaccaatttcagagagaacaatgtctgcctg	660
Db	601	AGCCTCCGCGGGAATACTAGTAACTGCTTCAACCAATTTACAGAGACAATGCTCGCTCG	660
OY	661	tctctgaatgtgcagtcacccaatgtgtggcagaagtgtcgtggagaacaagaacagatgag	720
Db	661	TCTCTGAGTGCAGTCAACCCAATGTTGGGCAATGTGCGAATAAACAAGATGATAG	720
OY	721	gatgtgacagaagaagataacagtcgttcttgaacctgttggacatgtctgcacacaggtttg	780
Db	721	GATGTGACAGAAAGAGATTAACAGTGTGTGTGAACCTGTTGGACATGCTGACACGGGTTTG	780
OY	781	gagcatatacccaacttctctctgtgatacatgtgtaaagctcgttagaagttcccaagat	840
Db	781	GAGCATATATCCCAACTTCTCTGATGATATGTAAAGCTGTGAAGTCCCAACAT	840

QY	841	gaggggctctgggaatccatctgtagtgctctttagtgcttcgggcgcggaacccctggg	900
Db	841	GGAGGGCCTCGGGAAATCCATGTAGTGCCTTTCAGTGCAGGGGGCGAATCCTTGGGG	900
QY	901	ttatttagtaaaacagatttggaagaaggctggtaaagctcgaaacatgaaatccttttcgttag	960
Db	901	TTATTATTGTAATAACGATTGGAGAAAGGTGGTAAGCTGAAACATGAAAATCTTTTTCGTAG	960
QY	961	aatgattgcattgtcaggaaltatgatlbgcgaccttcgaalatagaaalttlyaaacaaagca	1020
Db	961	AATGATTGCATTGTGACGATTTAATGATGGGACCTTCGAATAAGTAAGATTTCACAAACA	1020
QY	1021	caacatagttttcgcaagccatgctgtacccatcaatttgttccatgltccatgttcctgcga	1080
Db	1021	CAACATTATGTTTCCCAAGGACATGCGTACACCCATCATTTGGTTCCATGTGGTTCCTGCA	1080
QY	1081	gcaaataaaggcagatctgaacaactatcccaagtgagaagaacaatatcatcttcaagc	1140
Db	1081	GCAAAATAAAGCAGTATGAAACAATATCCCAAGTAGAAGAGACAAATTAATTCAACG	1140
QY	1141	cgltttagcccttgcagccagcatatattgaaacaagagtgtaacagltgcagcgcttccac	1200
Db	1141	CGTTTAGCCCTTGACACCGATATATGACAAGAGTGTGAACAGTGCACGCGCTTAC	1200
QY	1201	acggtgcagagagaccccgcatgaaccacgcctgtgacagatagatctcaactcaaga	1260
Db	1201	ACGGGCGAGAGAGACCCCGCATGAACACCCGCTGACAGATGACATCTCATCAAGA	1260
QY	1261	ctacctatagcgacacccctcgggaaaaacacatctcgctccagccttcggaaccttag	1320
Db	1261	CTACCTCTATACCGACACCCCTCGGGAAAAACACATCGCTCCACGCTCGGACACTTAG	1320
QY	1321	aatgattttagtaagctgtgaagcagtgtgtataacacaaataaataagcaagagctt	1380
Db	1321	AATGATTATTACTACGACTGTATAGCAGTGGTTATTAACCAAAAAAATTAGCAAGAGCTT	1380
QY	1381	aataaccagcttaagaaaggtacagaaggttlyggaattcagatcaccttccagagatgta	1440
Db	1381	AATATCCAGCTTAGAAAGGTAGACGAAGGTGGATTGGATCCAGATCACTTCCAGAGATGTA	1440
QY	1441	acaaataggttgctcaagctccaactatgtgaaaaacaattctcccgggggcgcgcaat	1500
Db	1441	ACAAATAGGTGGCTCAGCTCCAAATCTATGTAAAAAATCTCCCGGGGGCGGCCCAAT	1500
QY	1501	caggatgvcgaactaaaggcagagacacactatagaagtgaaatlygagtagatttagtg	1560
Db	1501	CAGGATGGCCCACTTAAAGCAGAGACAGACTTATATGAGAGTAAATGAGATCAATTTAGTG	1560
QY	1561	ggcaaatcccaagaagaagttgttctgcgttltagaagcacaagaatlggaagaaactgtg	1620
Db	1561	GGCAAAATCCCAAGAGGAAGTTGTTGCTGTGTAGAAGCACCAAGATGGAAGAACTGTG	1620
QY	1621	agccttctgttcttctgcgcaggaagacgccttccacccaaggaactgaaatgcagaacca	1680
Db	1621	AGCCTTCTGCTCTTTCGCCACGGAAGACGCTTCACCCAAAGGAAGCT-----	1667
QY	1681	agccagatgcagatltccaazaagaacgaagcagaagatgagagatatgttcttcaact	1740
Db	1681	AGCCAGATGCAGATLTCCAAGAAAGCAAGAGATGAGATATGTCTTTTACACT	1701
QY	1741	gatggcaccagaagaattcttgcacattgaaagtcaccaactaatgatactcagatctgcagc	1800
Db	1702	GATGGCACCAAGGAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGAGATCTCAGCG	1761
QY	1801	cttgggtgcagtgccaaggttaaccggtcaaaaagagaaaccaagcagatttggaaatcttt	1860
Db	1762	CTTGGTGTCACTGTCAAAAGGATACCGGTCAAAAGAAACACAGCAGATTTGGGAATCTTT	1821
QY	1861	gtcaagttccatattatgtgagagcagacatctaaagatlyggaagcttcgggttgaatgat	1920
Db	1822	GTCNAAGTCAATTATTAATGCGAGACAGCACTTAAGATGGAAGGCTTCGGGTGANTGAT	1881
QY	1921	caactgcagcagatgaaatlyggaatcccttltggtgcagatgacaacaacagaatgcccattgaa	1980

```

Db 1882 CAACCTGATTACAGCTAAATGAGAAATCCCTGTTGGCCAAAGACAAACCAAGATGCCATGGAA 1941
QY 1981 accctaaagaaggtctatgtctacttaagaagcaataaagaagaatgatccactatgtt 2040
Db 1942 ACCCTAAGAAGGCTATGTCTATCTGTAAGGCAATTAACACAGAAATGATCCAGCTTATTTGTT 2001
QY 2041 gcaagaagaataaagaagaatgcaatgagtcgaatgcacttgagagcccccctgagacctgag 2100
Db 2002 GCAGAGGAAATTAAGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 2061
QY 2101 ctgcccatgaaacagcgttgatgatagagaagaagaattccactccctcactcagt 2160
Db 2062 CTGCCCATTTGAACAGCCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2121
QY 2161 gggatggagggcttgatgataatgcacagagaatgagcctgagcctgagcctgagcctgagcct 2220
Db 2122 GGGATTTGGGGGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2178
QY 2221 gattcaggttaataaccagctgtccctcactcagtgataatgcccagaatgagcactgact 2280
Db 2179 -----GATTAATACCAAGCTGCTCCCTACAGATGATATCCCAAGATGATGATGATGATGAT 2232
QY 2281 atagaagatacagaggtgtccagatgtctccctcactcactcactcactcactcactcactcactc 2340
Db 2233 ATAGAAGATGACAGGTTGGCCAGTCCCTCCACATCTCTGACCAATGCTCTTCCAGC 2292
QY 2341 tcccatagatgatgtggggttgatgacgagatgtgtgacttgagcagaagctgcacatc 2400
Db 2293 TCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2352
QY 2401 agtgatcagcagcagctgtctcttgatgacagatgtgtatcactcactcactcactcactcact 2460
Db 2353 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2412
QY 2461 gaaagatctgagcagcagatgtgtcagaagaacagcaagaacatctcagaatgagcagc 2520
Db 2413 GAAGGATTTGGACGTCAG----- 2430
QY 2521 caatgtgatttcgttaaaacagaaatcaaaagatgattagttagttagttagttagttagtt 2580
Db 2431 ----- 2442
QY 2581 actaaactcaatacagtgatgacccaagaagcaggtctcccaagaagaatgtgtgtct 2640
Db 2443 ACTAACTCAATACAGTGTGATGACAGAAAGCAGTCTCCACAGCAGATGTGGTCT 2502
QY 2641 tccctggtctgagaagaatgaatcgtgtgagagctgtgacagcagcagcagcagcagcagcag 2700
Db 2503 TCCCTGGTCTGAGAAAGTCAAGCTCATTTGAGAGTCTGACAGCCGAGTTCGCCAGGTG 2562
QY 2701 acttgaatggggatattcccttccatcgtcccaagcagcagcagcagcagcagcagcagcag 2760
Db 2563 ACTTTGAATGGGATATTCCTTCCATGCTCCACGCCGCCGATTAATCAAGGAGGGA 2622
QY 2761 tgaatgagagctcagagctgtccatgacaaatctatgataaaccagcagcagcagcagcag 2820
Db 2623 TGCATGAGAGCTTGAAGCTGCTCATGACCAATCTTATATTAACCCCGGTGATGAT 2682
QY 2821 gattgtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2880
Db 2683 GATGATGAGAGCTTGAAGCTTGAAGAGACACAGAAAGATTAAGATTAAGATTAAGATTAAGAT 2742
QY 2881 gattcgtatccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2940
Db 2743 GAGTCTGATTCACAGCAGCTGATCAAGCTTCCACTCTCTGAGAGACAAATGATGATGATGATGAT 2802
QY 2941 aacaaagaagaagtgatgaactgtagaataaagaagaagaagaagaagaagaagaagaagaaga 3000
Db 2803 AACCAAGAAAGAGTGTATGAGACTGTAGAAAAGAGATTAAGTGGGAAAAGAAAGAG 2862
QY 3001 aagatagagataagagagagataaataagaaagcagaagaagagagagagagagagagagagag 3060
|||||
Db 2863 AAGATAGAGATTAAGAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2922
QY 3061 gaaagatggtcaggttgcaaaacatcgaagaatgatcagagatcagagcagcagcagcagcag 2922
Db 2923 GGAGACATGTT----- 2934
QY 3121 ataaatacagaagatccttaccatcagaagagagagagagagagagagagagagagagagag 3180
Db 2935 ----- 2934
QY 3181 gaaagatcagaagcagaactcagaatlaaggagaacagacgtcgaagcagcagcagcagcag 2934
Db 2935 -----AGATTTCAAGCAGAACTGAGAAATTAAGAGAAAGCAGACGCTCAGACGCTGAT 3240
QY 3241 gctgaatcagaatcttaccatcagaatcagcagcagcagcagcagcagcagcagcagcagcag 2991
Db 2992 GCTGAATTTCAAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
QY 3301 gttcctctatgaaggttccatgagcgtcccaagcagcagcagcagcagcagcagcagcagcag 3051
Db 3052 GTTCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3111
QY 3361 catatgatgatgtgtgtatgacagcagcagcagcagcagcagcagcagcagcagcagcagcag 3111
Db 3112 CATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3420
QY 3421 gtagacagtaacagatcactcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3171
Db 3172 GTAGACAGTAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3231
QY 3481 tttaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3231
Db 3232 TTTAGAGAAAGAAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3540
QY 3541 caacccctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3291
Db 3292 CAACCTGGCCGAAAGCAGAGCGCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
QY 3601 gtagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3351
Db 3352 GTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3660
QY 3661 taccagctctgctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3411
Db 3412 TACAGCTCTCTGCTCGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 3720
QY 3721 gtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3471
Db 3472 GAGCAGAACTTCTCCCTGGGGAAGGCTTCACAGATGCTCAAGATGCTCAAGATGCTCAAGATGCT 3780
QY 3781 agctacaaagcgtccagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3531
Db 3532 AGCTTACCAAGGCTCCAGGAAGGCTTACCTGAGAGCAAGAGGCTTCAAGGCTTCAAGGCTTCA 3840
QY 3841 ctggaactcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3591
Db 3592 CTGGAATCTAGAGAGCTCTCTCCAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
QY 3901 cagcctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3651
Db 3652 CAAGCTCTCTCCAGAGGAGGCGCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 3960
QY 3961 taagccctcccaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3711
Db 3712 TACGCCCTCTCCAAAGGAGGCTTCCGAGCAAGATGATGATGATGATGATGATGATGATGATGATGAT 4020
QY 4021 agcgtgaacagactcagactcctgagaagaagagagagagagagagagagagagagagagag 3771
Db 3772 AGGCTGAACAGACTTCAAGATCTGAGAAAGAGAGGCTTCTATTTCTGGA 3822
|||||
RESULT 7
AF252293

```

LOCUS	AF252293	3801 bp	mRNA	linear	PRI 10-JAN-2002
DEFINITION	Homo sapiens partition-defective 3 splice variant c (PAR3) mRNA, complete cds, alternatively spliced.				
ACCESSION	AF252293				
VERSION	AF252293.1	GI:8037914			
KEYWORDS	human.				
SOURCE					
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 3801)				
TITLE	Joberthy,G., Petersen,C., Gao,L. and Macara,I.G.				
JOURNAL	The cell-polarity protein Par6 links Par3 and atypical protein kinase C to Cdc42				
MEDLINE	Nat. Cell Biol. 2 (8), 531-539 (2000)				
PUBMED	20394296				
REFERENCE	10934474.				
AUTHORS	2 (bases 1 to 3801)				
TITLE	Gao,L., Macara,I.G. and Joberthy,G.				
JOURNAL	Multiple splice variants of Par3 and of a novel related gene, Par3L, produce functionally different proteins				
REFERENCE	Unpublished				
AUTHORS	3 (bases 1 to 3801)				
TITLE	Joberthy,G.M., Petersen,C.D., Gao,L. and Macara,I.G.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (05-APR-2000) Markey Center for Cell Signaling,				
FEATURES	University of Virginia, West Complex, P.O. Box 800577,				
SOURCE	Charlottesville, VA 22908-0577, USA				
	Location/Qualifiers				
	1..3801				

QY 961 aatgatgcatlgtcaagattatgaatgagacccctcgaaatagaagatttgaacaagca 1020
Db 829 AATGATGTCATGTCAGAGTTATATATGCGACCTTCCAAATPAGAAATTTGAACAGCA 888
QY 1021 caacatattttccccaagcaatcgtaacccatcatlgtgttcagttgttctcgca 1080
Db 889 CAACATATATTTTTCGCCAAGCCATGCGTACACCCATCTTTGGTTCCATGTGGTCTCGCA 948
QY 1081 gcaaataaagcaglatgaacaactatcccaagtgagaagaacttaactaactaagc 1140
Db 949 GCAAATTAAGAGCAGTATGAAACAATCTCCAAAGTGAAGAAACATTAATCTATCTCAAG 1008
QY 1141 cgtttagccctgagacccagctatatattgacaacaggagtgagaagtgacagtgcaac 1200
Db 1009 CGTTTACCCCTGACAGCCAGTATATATGACACAGAGTGTGAACAGTCCAGGGCTTCAC 1068
QY 1201 agcgtcagagagaccccgacagaccccgctgaacagatagactctcaactaaga 1260
Db 1069 ACGGTGAGAGAGACCCCGACAGCAACCCGCTGACAGCATATGACTCTCACTCAAGA 1128
QY 1261 ctaccatagcgacacccctcgagaaacacacacacacacacacacacacacacacac 1320
Db 1129 CTACCTCATAGCCACACCCCTCGGGAAAAACCATCCGCTCCAGCTCGGCACTCAG 1188
QY 1321 aatgtatttagtaagactgtlaagcagtggtlaaacacacacacacacacacacac 1380
Db 1189 AATGTATTTAGTACGACTGTAGCAGTGTATTAACACCAAAATTAAGCAAGGCTT 1248
QY 1381 aatatacagacttaagaagaagatagaagtttggatcagacacacacacacacacac 1440
Db 1249 AATATACACTTAAACAAAGTACAAAGTGTGGATTCAGACTCACTTCCAGAGATGTA 1308
QY 1441 acaatagtggtcagctcccaatcatalgtgaanaacacacacacacacacacacac 1500
Db 1309 ACAATAGGTGGCTCAGCTCAATCTATGTGAAAAAATTTCCCGGGGGGGGCGCAT 1368
QY 1501 caagatgcccagacttaagcagagacagacactatagaagtaaaatgagtagaattag 1560
Db 1369 CAGGATGGCCGACTTAAGCGAGAGACAGACTTATAGAGTAAATGAGTAGATTAGTG 1428
QY 1561 ggcacatcccaagaaggaagtgttgcgtgtlgaagaagcacaagaatgaaagaactgt 1620
Db 1429 GGCAAATCCCAAGAGGAAGTGTGCTGTGTGAGAAAGCAACAAATGGAGAACTGTG 1488
QY 1621 agccttctgtcttcgcccagagagacgcttccacccaagggaacttgaatgcagagca 1680
Db 1489 AGCCTTCTGTGTTTGGCCAGGAAGACGCTTCACCCAAAGGGAAT----- 1535
QY 1681 agccagatgcagattccaaaagaagaagaatgagatattgttcttaacact 1740
Db 1536 -----GAAAGCAGAAAGATGAGGATATTTGTTCTTAACACT 1599
QY 1741 gatgacacaggaatttctgacatttgaagtcccaacttaattgaatcagatctgcagc 1800
Db 1570 GATGGACCAAGGAATTTCTGACATTTAAAGTCCCACTTAATGATTCAGATCTGCAGGC 1629
QY 1801 ctgtgttcacagtgtaaaaggtataccogtcaaaagaagaacacagagattggaaacttt 1860
Db 1630 CTTGCTGTCAAGTCAAAAGGTACCGGTCAAAAGAAACCAACAGATTTTGGAAATCTTT 1689
QY 1861 gtcaagttcattatlaatgagagagacacactaaagatggaagcttcgggttgaatga 1920
Db 1690 GTCAAGTTCATTTAATGAGAGAGACGATCTTAAGATGCAAGGCTTCGGGGATGAT 1749
QY 1921 caactgataagaataatgagaatccctgtlgtgcaagaacacacacacacacacacac 1980
Db 1750 CAACTGATAGCAGTAATGAGATCCCTGTTGGCAAGAACAAACAGATCCCATGTGAA 1809
QY 1981 accctaaagaaggtctatgtctactgaagcaataacagaggaatgataccagcttaatt 2040
Db 1810 ACCCTAAGAGGTCTATGTCTACTGAAGCAATTAACGAGGAATGATCCACTATTGTT 1869

QY 2041 gcaagagaataaagcaagtcgaatgaagctgaagtcacacttggaagccccctgacactgag 2100
Db 1870 GCAAGGACAAATAAACCAAGTGCATAGAGCTGAAATCTGAGGAGCCCCCTGAGACTGAG 1929
QY 2101 ctgcccattgaanaagcgttggatgataagaaacagaaatcccatccctcaag 2160
Db 1930 CTGCCATTAACAAAGCGTTGATGATAGAGAAACGAAAGATTTCCCATTTCCCTACAGT 1989
QY 2161 gggatcaggggcttgatgaatcgccacagaaatgctgccctcaagataagatggt 2220
Db 1990 GCGATTAAGGGCTTGTATGATACCCAAAGCAAAATGCTGCCCTCAGTAAAGATGATG 2046
QY 2221 gagtcaggttaataacagacttccctcaagtgaaatgcccacaaatgaaactgtact 2280
Db 2047 -----GATTAATACCAAGCTGTCCCTACAGTAATATGCCCAAGATGACATGTAT 2100
QY 2281 atagaagatgacaagttgttcagtgcttcccaacacacacacacacacacacacac 2340
Db 2101 ATAGAAGATGACAGGTTGCGAGTGTCTCTCCACATCTCTGACACAGTCTCTTCAG 2160
QY 2341 tcccatgatatgtgtgggttgtgacgagatgtgtacttgggccaagtgtaac 2400
Db 2161 TCCCATGATGATGTGGGTTGTGACGCGAGATGCTGTGTAATGGGCAAGGCTCAATC 2220
QY 2401 agtgaatcagcagactgtcttctgagtcagagatgtatcagcttcttcaaga 2460
Db 2221 AGTGAATCAGCCGACTGCTTTGAGTCCAGATGTTGATTCAGATTCCTTTTCACGA 2280
QY 2461 gaagatttggagtcagatgatagtcagaanaaacgcacaagaactttcagatgcag 2520
Db 2281 GAAGATTGTGAGCTCAG----- 2298
QY 2521 caattgattcgttaaaacagaaatacaaaagcatgattgatatgctgaag 2580
Db 2299 -----ATAGCTGACGAG 2310
QY 2581 actaaactcaatacagtgatgaacagaaagcagttctccacagagatggtgtct 2640
Db 2311 ACTAAATCAATACATGATGATGATGACAGAAAGCAGGTTCTCCACAGAGATGTGGTCT 2370
QY 2641 tccctgtgtctgaagaagtcagtcgttggagagtcgcagacccgaagtcgcagagt 2700
Db 2371 TCCCTGTGTCTGAAGAAGCAAGTCTGTGAGAGTCTGACAGACCGAGTGTGCCGAGTG 2430
QY 2701 acttgaatggagatattcccttccatctgtccagcgcgcgcgcgcgcgcgcgcgcgc 2760
Db 2431 ACTTGAATGGAGATATTTCTTTCATCTGTCACGCGCGGATTAATCAGAGCCAGGGA 2490
QY 2761 tgcattgagagcttcaagatgagctgcacacacacacacacacacacacacacacac 2820
Db 2491 TGCATGAGAGCTTCAGAGCTGCCATGACCAATCTTATGATTAACCCCGGTAGATGAT 2550
QY 2821 gatgatgaagcagtcagagacttggaaagaacacagaaagaatcagatcagagagga 2880
Db 2551 GATGATGAGGCGCATGAGACCTTGAAGAAACACAGAGAAAGTTCAAGATCAGGAGGA 2610
QY 2881 gagctgtatccacagcagtgatcagcttcccaactctctgaaagacaaatgaatga 2940
Db 2611 GAGCTGTATCCACAGCGAGTATCAGCTTCCCACTCTCTGAGACGACCAATATGAGA 2670
QY 2941 aaccagaagaagtgataagactgataaanaaagaatlaaactgynaagaagaagaag 3000
Db 2671 AACCAAGAGAAAGCTATAGACTGTAGAAAAGGATTAACCTGGAATAAAGAAAG 2730
QY 3001 aaagataagataaagagaagataaataagaagccaagaaggaatgctgaagggcttg 3060
Db 2731 AAAGATAGAGATTAAGAGAGATTAATTAAGAACCAAGAGAGGAAATGTGAAGGCTTG 2790
QY 3061 ggaagacatltcaggttggcaaacatctgaaagaatgaagaatgagaagaaggggttaa 3120
Db 2791 GGAGACATGTTGAGTTTGGCAAAACATCGAAAGATACCAAGATTCAGAAACCGCTTAA 2850
QY 3121 ataaaaatacagaatcccttatacagaagagagagatgaatgaagcagagagcag 3180

[illegible][illegible]

RESULT	8
AF467005	
LOCUS	3735 bp mRNA linear PRI 07-FEB-2007
DEFINITION	Homo sapiens partitioning defective 3 protein splice variant e
	(PRND) mRNA, complete cds, alternatively spliced.
ACCESSION	AF467005
VERSION	AF467005.1 GI:18568349
KEYWORDS	
SOURCE	human.

Query Match	Similarity	58.8%	Score 3347.4	DB 9	Length 3735
Best Local	Similarity	91.7%	Pred. NO. 0		
Matches 3734	Conservative	0	Mismatches	1	Indels 336
					Gaps 5
Oy	1 atgaagaatgacccgtgtgcttcgagacggaaccgggtgtgtcgtgtgcgttgcggagacggtccac	60			
Db	1 ATGAAGAATGACCCGTGTGCTTCGAGACGCCGCCGGGAGTGCTGTGCCCTTGGGGGAGACGCCAC	60			
Oy	61 atgaagaatttcagcctcatccagcaagcgtgtgacccgtctacccggaagaagccatctgccaag	120			

OY	2281	atgagaatgaagaagcttgcagagcttcctccaacatcctcagacgaatcctctccaagc	2340
Db	2101	ATGAAATATACAGGTTGGCCAGGCTTCTCCCAACATCTCTCGAACATCTCTTCCAGC	2160
OY	2341	tcacatataatcttgaggttcttgacacgcaga tgcctggtactcttgccaaagctctgaac	2400
Db	2161	TCCCATATATATGTGGGGTTTGTGACGGCAAGATGCTGTGACTGTGGCCAAAGCTGCATC	2220
OY	2401	agtgatccaagccagctgcctcttgaagtcagaatgltgaatccaagtcctgtcttccaaga	2460
Db	2221	AGTGATTTACACCCACATGCTCTTGTGAATGCCAGATGTGTATCCAGTTCTGCTTTCACAGA	2280
OY	2521	caatctgattctctgttaaaacacgaanaatcaaaagaatgaatgaatgaatgaatgaatgaatga	2580
Db	2341	CAATGTGATTTCTGTTAAACACGAAATATAAAAGCATGATTTAGGTA-----	2399
OY	2581	actaaactcaatacagctggaatgacacgaagaagcttctccaagcagaagatgctgctct	2640
Db	2390	-----GTTCTCCAGCAGAGATGTGGGTCT	2415
OY	2641	tcctctggtctgaagaagltcaagctcgcttggaagctctgcagaccgcagcttgccagagtg	2700
Db	2416	TCCCTGGGTGTGAAGAGTCAACCTGTTGGAGAGCTGTGCAACCGCAGTTGCGGAGGTG	2475
OY	2701	actctgaatgggaatctcctcttccatctgtccaagccgcgcgaataataagaagcagaagga	2760
Db	2476	ACTTTGAATGGGATATTCTTTCATCTGTCACAGGCCCGGATTAATGAGGCGAGGGGA	2535
OY	2761	tgcaatgaagctctcagagctgcgcacgcgaacatcttatgataaacccgcggtagatgat	2820
Db	2536	TGCAATGAGAGCTTCAGAGCTGCCATTCGACAAATCTTATGATTAACCCGGGTGATGAT	2595
OY	2821	gataatgaagagctatgtagagacctctggaagaagacacagaagaagltcaagaatcaagga	2880
Db	2596	GATGATTAAGCGCATGCGAGACCTTGGGAAGACACAGGAAGAAAGTTCAAGATCACGAGGA	2655
OY	2881	gagctgtataccaagaacag tga tcaagctctccaactctctgtagaagaacaatgaatga	2940
Db	2656	GAGCTGTATCCACAGCGCATGTATCAGCCTTCCACCTCTCTGTGAAGACAAATGATGGA	2715
OY	2941	aaccaagaagaaggttga taaagactga tagaanaaagyataaaactctggaagaagaagaag	3000
Db	2716	AACCAAGAGAAAGTGATTAAGACTGATAGAAAAAGATTAAGATTAAGATTAAGATTAAG	2775
OY	3001	aaagataagataaagagaagaagataaaatgaagaaccaaagaaggaatctgaagagcttg	3060
Db	2776	AAAGATATGAGATTAAGAGAGATTAAGATTAAGAAAGCCCAAGAAAGGAATCTGAAAGGCTTGG	2835
OY	3061	ggagaacatgttccaggttctgccaacatcgaaaaagatgaagaattggaagaaaaaggtataa	3120
Db	2836	GGAGACATGTTC-----	2847
OY	3121	ataaaaatacagatctccttatcatcagaagagagagatacgatgaagcagagcag	3180
Db	2848	-----	2847
OY	3181	gagagatccaagccaaactcgaagaattaggaacgacaaagctcgaagagcgtgactat	3240
Db	2848	---AGGATTTCAAGCCAAACCTCGAGAAATTTAGGGAAGCAGCAAGCTCAGAGCGGTGACTAT	2904
OY	3241	gctgaatatcaagatttccatcgggaacttggctgtgatagataagttaatgtatagtgagga	3300
Db	2905	GCTGAATTTCAAGATTTTTCATCGACATTTTGCTGTGATGATGATGATTAATGATTTGGGGGA	2964
OY	3301	gttctctcttgaaggttccaatgggcctcaaaagcttaagacctcaagagcccaagcagaagaag	3360
Db	2965	GTTTCTTTTATGAAAGGTTCCATGAGCTCTCAAGCTTAGACCTCAGACCCACGAGAAAGGG	3022

[illegible]

Db	1700	ACTTATAGGTGACTAGCTCCCAATTTATGTGACAAATTTCTTCTTCGAGGGCTGCACATT	1759
OY	1501	caggatgagccacattaaagcagagaaacaactatagaagtaaatggaatggaatttaag	1560
Db	1760	CAGATGTGGCAGGCTCAAGGGGGGAGACCGGCTAATAGAGGTCAATGAGAGTATGATTTTACGG	1819
OY	1561	ggcaaatcccaagagaagattgttcctgctgtgttgagaagcacaagaatggaagaactgtg	1620
Db	1820	GGCAAGTCCCGAGGAGAAAGTTGTTTCCCTGTGTGAGAAAGCACAAAGATGGAGGGACGGTG	1879
OY	1621	agccttcgtgtcttcgcgcagagaagacgcctccaccagaaggaactgaatgcagagcca	1680
Db	1880	AGCCCTTGTGGTTCCTTGCCTCAAGAAAGGGCTTTCACCCCAAGGGAAATGAAATGCCAATCCC	1939
OY	1681	agccagaatgcagattcccaaaagaacgaaagcagaagaatggaagatattgttcttaacct	1740
Db	1940	AGCCAGATGCAGAGTCCAAAAGAAAGAAAGCCGAAGACGAGAGCAATGTCTCACACTT	1999
OY	1741	gatggcaccagaggaattctctgacatttgaagtcacctaatagtattcagaatctgcgaagc	1800
Db	2000	GACGGTACCAAGGAGATTCTTGACTTTTCGAAGTTCCACTGAATGACTCAGAGGCTTCACAGG	2059
OY	1801	cttggtgtcagtggtccaagaagttaaccggtccaagaagaaacccagcagatttgggaactctt	1860
Db	2060	CTTGTGTGCACCGTCGAAGGGGGAACCGCTCCAAAGAAACCAATGCAAGATTTGGGATCTTTC	2119
OY	1861	gtcaagtcacattatataatgagagagcagcatcctaaagaatggaaggtccggatgaatgat	1920
Db	2120	GTTAAGTCCATCATTTAAACGGGGGAGCTGCATCTAAAGATGGGAGGCTCAGGGTAAATATAC	2179
OY	1921	caactgtatagcagtaaaatggaagataccctgtgttgycgaagacaacaagaatgtccatgga	1980
Db	2180	CAGCTGTATAGCTGTGAACGGAGAGTCTGTGCTGGCAAAAGCCAAACCAAGAAAGCCATGGAG	2239
OY	1981	accctaagaaggtcttatgtctactcgtaagcgaataaacaagagaatgatccagtatgtgt	2040
Db	2240	ACTGTGAGGAGGTCCATGTCCACCGAGGGCAACAAACCTGTGCATGATCAGCTCATCTCGTG	2299
OY	2041	gcaagaggaataaagaagaatgcaatggagctggaagtacccctgggaagcccccctggaacctgag	2100
Db	2300	GGGAGGGGAGTCAACAAATATTAACGAGATGGCGGTCTCTCTGGAGCCCGCTGCACCCAG	2359
OY	2101	ctgcaccattgaaaacggtcttgaatgaatagaagacgaagaatttccattccctctacagt	2160
Db	2360	CTGGCCATAGAAACAGAGTGTGATGACAGAGAAAGCAGGATCTCACACTCCCTCTACAGT	2419
OY	2161	gggatctgaaggggtctgaatgcgcgccagcagaagaatgtgcctcagtagaataatgggt	2220
Db	2420	GGGATCGAGGGGCTGATGTGATGCTCCACACAGGAATGCCGCACTACAGAGCATATGAGGT	2479
OY	2221	gaagtcaggtaaataacaaagcgtctcccttaagaatgaatatacccccaagaatgaaacgttaatt	2280
Db	2480	GAGTCAAGTAAATATCCAGGCTCCGCCAACCGGTGAATATCCACAAATGATGACACTGTCTATG	2539
OY	2281	atagaagaatgcagaggtgtgcacagtgctctccacatctctcctgaacagtgctcctctccagc	2340
Db	2540	ATTGAAGATGACAGAGTGCCTGTGTCTCCCTCCACCTCTTGACACAGTCCCTCCACAGC	2599
OY	2341	tcaccaatgaatgtggggtttgtgacgagcaaatgtctgtaacttgggccaaggtctgaact	2400
Db	2600	TCCCATGTATGACGTGGGATTCATTAATGACAGAAAGCAGGCACATGGGGCAAGGCTACCATC	2659
OY	2401	agtatattcagccagcgtctctcttgaatccaagaatggtgacaggtctctgtttcttaaga	2460
Db	2660	AGTGCATCAGCCGACTCTCTATTGAGTCCAGATGTGATCCGGTCTCGCTTTTCAACGA	2719
OY	2461	gaaggaatttgacgtcagagtatgtcagaanaaacgcacaagaacaaatttcagatgcaggt	2520
Db	2720	GAAAGATTGTGAGCCGAGAGTATGTCAAGAAAAAGCACAAGCAATTTTCAATGCCAGT	2779
OY	2521	caattggatttcgtttaaacaacgaaatcnaaaagcaygaattgagtatagctgaagag	2580

```

OY 3661 taagctctctgctcgcaagcaggaaatgacgctcggtctccaggaactcttg 3720
Db 3914 TACAGCTCCCTCCAGAGCAAGAGAGAAATCCAGCTCTATCAAGGATTCCTCG 3973
OY 3721 gagcagaactactcccttgaggaggtctccagagtgcagaagaagaccaggtactcc 3780
Db 3974 GAACAGAACTAGGCTCCCTGGTGAAGGCTTCCAGAGTGCCAGAGAAACCCAGGATATTCG 4033
OY 3781 agctacaagagctccaggaagcagctactcgtgagagagcagcttcaagcagggtcatg 3840
Db 4034 AGTTACAGAGGCTCCAGAGAGCGCTATCAAGTGGCCATGCTTTAAATCCAGGCTCATG 4093
OY 3841 ctggaactaaggaggtctctcgccaggaagcaggagcaggagcagcagatggaag 3900
Db 4094 CTGAGAGCCAGAGAGCTTCTCCGACAGGAGCAGAGCGCAAGAGACAGCACTGAGAGAG 4153
OY 3901 cagctctctcagaggagggccagcaactatgctctgataagaaggtccagagcccaagt 3960
Db 4154 CAGCCACAGCGAGTGGATGAGTCAAG----- 4177
OY 3961 taagcccccagaggggcccttcggaagatgagcccccctccctctcagatgtcg 4020
Db 4178 -----GGGCCCCCTCCGAGAGATGATACCCCGCTCCATGTCAGAGTGCC 4222
OY 4021 aagctgagacagactcagactcctgagaagaggagccctctctactcagc---acgc 4077
Db 4223 AGCTGAACAGACTGACAGACCCGAGAAAGGCGCCCTTCTACTCTTAGGCTCAATG 4282
OY 4078 aataacgagatgctcagctgagcgaataaagaacattctcctatgagaagctgtatct 4137
Db 4283 AGGATGAGAGCATTTTGGTGTATCAATTAAGAGCATTTCCCGTGAGAGCTGTCTCCTC 4342
OY 4138 gggaggtttttaaacc-----tcgagtactatgaggatcttctgttct 4185
Db 4343 TGGAGTTATTTTAAACAAAACCCCTCATTTGTGTGTACTACAGAACCTTCTGTCACT 4402
OY 4186 ggtatcagctgcttgaagcgtgtagcagaagaatggaagcgttaagtcttctgcac 4245
Db 4403 GGTATCAGTGTCTTAAAGCAGAGAGCAAAAGAAAGGAGGCTTAATGTCTTGCCAC 4462
OY 4246 tatgtctcaagtctcgttctcagtagaagattcccaaccgtgacatcatctgttga 4305
Db 4463 -----GCCTATATCTCGTGAAGAGACCCCATCCCAACCCCAACCTCAAGT 4512
OY 4306 ggttcaatagctctgagccctcctcaagatcaggaattcggccctactcatgagtt 4365
Db 4513 TCAAGGACATCTACTCCCTCTCTCAAGCATCCGAAACTTCACCTCATATTAAG----- 4568
OY 4366 gtccgagcttggtgtaaacatccctgacacactgacgaatcaactcagtaagtgg 4425
Db 4569 -----GTGCCGTGTGAGACGTACAGACTTCATCTTAAGAG 4602
OY 4426 aggatgacgtgagccctgttctccctcctaagttctcttagctatgga--tgacatct 4483
Db 4603 GGGACCCCCCCCCCCCCCAGCGTGGCCGACACCCGTTCCCTGCGAGCGTACATCT 4662
OY 4484 tagtctctgtaggaagaatgaggcagatcacacaaatattgggagcttctgtaact 4543
Db 4663 CGGCTCTCGAGAGAGGAAGTGCTGACAGATGACAGAGGAGGAGGCTTTCCAGAACTT 4722
OY 4544 cacaagaagcatttgcgtacttctcactgacactggtgttcttctctctctcctcagc 4603
Db 4723 CGCAGACAGGCTTACCATGTT--CTCCCACTCGGTCTGGGGTTCCTTTCTTCTTCA 4780
OY 4604 tcttttgaag--ggaagatgagcatctatcagaagactaagacatgctcacaatg 4662
Db 4781 TCTCGAGAGCGTGAAGGCTTAGGGTGTGATGAAGAACTTAAGCCATGAGCGCACTGCG 4840
OY 4663 tttagtcaac-----atgggggtctcttctgtaactgactcctatgacaaataacaa 4718
Db 4841 TTTGGTCTCTGGGGGTGAGGGTAACTTTCAACTGCTTATTAACCTCAC--ACCATAT 4898

```

OY 4719 agtattgatctctgctggttataacatacgttctg 4760
Db 4899 CACAGCATTTCTAGTCAGTGTATTAAGTACATGCGTTGAGT 4940

RESULT 10

AF196186 3470 bp mRNA linear PRI 16-NOV-2001
LOCUS AF196186
DEFINITION Homo sapiens atypical PKC isoform-specific interacting protein
ACCESSION AF196186
VERSION AF196186.1 GI:13491611
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

BASE COUNT 1003 a 811 c 934 g 721 t 1 others
ORIGIN

Query Match 53.6%; Score 3050.6; DB 9; Length 3470;
Best Local Similarity 99.6%; Pred. No. 0;

FEATURES

source

CDS

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
288..3383
/note="HASPsa"
/codon_start=1
/product="atypical PKC isoform-specific interacting protein short variant"
/protein_id="AAK27892.1"
/db_xref="GI:13491612"

/translation="MKVYCFGRTRVYVPCGDHMKVFLIDQAVYRYKAKAPNY
MIOVHRLHEDDGIILDDILCDVADKRLVAVPEODPHRGGTSASSTGDSPE
IFGSELTNNVSAFOPYOATSELEVPSYLRNMPDHYRRSDPALIGLSTVSQSP
SSEPSKRNPTNMTAGTAKONTAGSPCTCKRDKDENTRSIPROTSMNSQFQDNA
RSLSASHPWGKLEKODDEGTEDNSRVEPVGHADTGLEHLPNFSLDWVLE
VPNDGGPIGTHVVPARGGRITGLVRLKEGGAHEHNTLFRENDICIVRIDGDLVE
RREPOHMRQMRKPTIIFWHVVPANKREYOLQSEKNNYSRSRSQVITDR
SVNSACILHVQAPRLNHPPEQIDSHSLRPHSAHSPGPPASAPASQNVSTVSSG
YNTKIGKRLNIQKKGTEGLGFSITSDVTIGSGAPLYVNNITPRGATIDGRLAG
DRILEVNGVDVLGKQSEEVVSLRSTKMGRTSLVPEQEDAFIRREINAPSQMOP
KETKADDEDIVLPDGTREFLEFVPLNDGAGIGSVKGNRSKKNHNDGILVKS
INGGASKDGRIRVNDOLIAVNGESLGTKNDQDAETLRSMSTEGNKGMIOLIVR
KIOSPTVVMPPDDVITLEDRLPVLPPLHSDSSSSSHDDVGVPTADACTWAKAIS
DSADSLSPDVPDVLAFQREGRGQSKSEKRTKQFSDASQDLPVKTAKSKSMDLGLAD
ERKLNTVDQKAGSPSRDVGSLGKSSLESLQTAVAVYTLNGDLPFHRPRTIR
GKCNESFRAALIDKSYDRPAVDDEGTELEPTDESSSSGREGSVASDQSHSLE
R⁺ROMNNOEGDKTDKDKTKGKKKKRDKMKMKMKMLKGLDMFLAKIPKPK

	Matches	3072;	Conservative	0;	Mismatches	4;	Indels	9;	Gaps	1;
QY	1	atgaagaagtcg	atgctgtc	tcgacgaagca	cccggtgc	gtgcgtgc	tcgacgaagtcg	caac	60	
Db	288	ATGAAAGTGA	CGTGTGCTTGG	AGAGGAGCA	CCGGGTGTG	GTGCGTGGG	GGGAGCGGCCAC		347	
QY	61	atgaagcttc	caagcctca	ccagcaggc	ggtgacgc	gctac	cggaaagccatgc	ccaag	120	
Db	348	ATGAAAGTTTC	CAGCCTCATC	AGCAGGCG	GGTACCC	CGCTACCGG	AAAGGCCATTCGCCAAG		407	
QY	121	gatccaactc	ctcgtgat	ccaagtcat	ctgtgttg	aaactlga	gagatggaggaatactaa	ctaac	180	
Db	408	GATCCAAACT	CTCGATACAG	GTGATGCTTGG	AACATGGAGATGG	AGGAATGACTATGAC			467	
QY	181	cttgatgcatt	ctcttgc	tgatga	caagca	ataaaga	caagactgtg	taagctgtt	240	
Db	468	CTTGATGACAT	TTCTTGTGTGAT	GTAGACACAG	ATTAAGACAG	CTGTGTA	CACTGTTTGTAT		527	
QY	241	gagcagagtc	caacatca	tcagagag	gtgatg	caacag	gtgcagtlcca	cggtgta	300	
Db	528	GAGCAGGATTC	ACATCATCAG	GAGGTGATGG	ACCAATGTCAC	AGTTCACACGGGTAC	CCAGAGAC		587	
QY	301	ccagagatatt	gtgtga	gtgagctt	ggca	ccaacaat	tgctcaagcctt	caagctta	360	
Db	588	CCAGGATATTT	TGGTAGTGAG	CGCTGGCAC	CAACAAATGTCTC	AGCCCTTCAGCGCTTAC	CA		647	
QY	361	gcaacaagtga	aatgtag	gtca	caacccttc	agcttc	cttgagcaaatatgc	ctcttat	420	
Db	648	GCAACAAATGA	ATTTGAGAGT	CACACCTTC	AGCTTCGAC	CAAAATATGCTCTT	CATGTT		707	
QY	421	cgaagcag	tagtga	cccacag	ctcta	atgtgc	ctc	caactctc	480	
Db	708	CGACGCAATGA	TGAGACCCAG	CTCTAATTTGG	CCCTCCATCTT	GTCTGACAGTATG	TAATTT		767	
QY	481	tcctctga	agagcctt	caagaa	aaatcc	caacag	ctgtg	tcaaca	540	
Db	768	TCCTCTGGA	AGAGCCCTTC	CAGGAAATATCC	CAACCTGTGCT	CAACACAGCCTGGCTTC	CTC		827	
QY	541	aagcgaac	ctactgc	gtgagag	tcctta	aaaccc	tcgcgc	gaagaaga	600	
Db	828	AAGCGAAC	CACTGCTGGG	AGTCTTAA	AACTCTCGAC	AGAGAAAGAT	TAATACTTAC	ACA	887	
QY	601	agcctccc	cgagata	ctagta	actg	gtctta	ccaattc	cagagag	660	
Db	888	AGCCCTCCG	GGGATAC	TAGTAACTAG	TGCTTAA	CAATTTCA	GAGAGAC	AATGCTCGCTG	947	
QY	661	tcctctga	gtgcag	tca	cccacaa	tgtgtgg	cgaa	gtgc	720	
Db	948	TCCTCTGAC	TGTGCAG	TACCCCAAT	GGTGGGCA	GTGGCTGG	GAATAAAC	CAAGAACGATGAG	1007	
QY	721	gatggaga	caagaaga	gaata	caacag	tcgtgt	gaacctgt	tgagac	780	
Db	1008	GATGGGAC	AGAAAGAGAT	TAACAGTGTG	TGAACTGT	TGGACATG	CTGACACGGG	TTTG	1067	
QY	781	gagcaata	ccaactt	ctctc	gtga	tga	tatgtg	aaagctg	840	
Db	1068	GAGCATAT	ACCACCAAT	TTTCTGTGAT	GTATGTTAA	AGCTGTGTA	AGATGCCCAAC	AGAT	1127	
QY	841	ggaagg	ccctc	tgga	atcca	tgtag	tgcctt	ca	900	
Db	1128	GGAGGCGCT	GTGGGAAT	CCATGTACTGT	GCCTTTCAC	TGCTCGAGCGG	CGCAACCTCGGGG		1187	
QY	901	ttatattga	aaaaag	atlg	gaga	aaggtg	gtaa	agctg	960	
Db	1188	TTATATTG	TAAANCGATT	GGAGAAAG	TGTAAAGCT	GTAGACAT	GTAAATACTTTT	GTGTG	1247	
QY	961	aatgat	tgcat	gtgca	agat	taata	tga	tgcgc	1020	
Db	1248	AATGATG	TCATTTGC	AGATTTA	TATGTATG	CGACCTTC	CAAAATAGAA	AGATTTA	1307	
QY	1021	caacata	gtcttc	gcgaag	caatgc	taca	cccacat	catl	1080	
Db	1308	CACATAT	ATGTTTCC	CAAGCCAT	GTGTCAC	CACTCATAT	TGTTGGTTC	CAATGTTG	1367	

OY	1081	gcaataaagagcagctatgtacaacatcatcccaagtfgaagaacaatctattccaagc	11400
Db	1368	GCAAAATTAAGAGCAGTGTGAACAACATATCCCAAAAGGAGAACAAACAAATTACTATTTCAAGC	14272
OY	1141	cgcttcagccctgcacacagtatattgcaacaagagcgtgtaaacgtgcaggtctaac	12007
Db	1428	CGTTTACCCCTGACACCCAGTATATTGAAACAGSAGTGTACACAGTGCAGCGCTTCAAC	14878
OY	1201	acggtgcagagagcaccgcagctgaaaccaccgccttgacagatagactctactcaaga	12606
Db	1488	ACGGTGCAGAGAGCAGCCCCCACTGAACACACCCCGCTGACACATGTAGACTCTCATCAAGA	15474
OY	1261	ctacctatagcgacacccctctggaaaaaacacacacacacgcctgcacagccctgcagctag	13220
Db	1548	CTACCTCATACCGCACACCCCTCGGAAAAACACATCCGCTGCAGCTCCGCGACCTTGAG	16077
OY	1321	aatgtattagtaagactgtgaacgactgtgtataaacaaccaaaaaaagacaaagagctt	13800
Db	1608	AATGATTTTACTAGACTGTATAGAGAGTGGTTATTAACACCAAAAAAATATAGCAAGGCGTT	16677
OY	1381	aatatccagcttaagaaagttacagaagtttgggattcaagcatcacttccagagatga	14400
Db	1668	AATATCCAGCTTAAAGAAAGGTATACAGAAAGTTTGGGATTCAGATCATCTTCCAGAGATGTA	17227
OY	1441	acaataagttgcctcagctccaatcatatgtbaaaaacatctcccccgggggcgagcat	15000
Db	1728	ACAAATAGTGGCTCAGCTGCCAATCTATGTATAAATAACATTTCTCCCGGGGGCGGCAATT	17878
OY	1501	caggatgcgcgaacttaagcagagacagacattataaggtaaatltgagtagattagtg	15600
Db	1788	CAGGATGGCCCACTTAAAGCGACGAGCAGACTTATAGAGTAAATGAGATTAATTAGT	18477
OY	1561	ggcaatcccaagaggaagtgtttctgcgtltgagaagacacaagatggaagaagcttg	16220
Db	1848	GGCAATCCCAAGGGAAGTGTGCTGTGTAGAGAACCCAGTAGTGAAGAACTGTG	19077
OY	1621	agcctctgtgcttctgcgcagagaagaagccttccaccaagagaaactgtatgcagagca	16800
Db	1908	AGCCTTGTGGCTTTCGCCCGAGGAAGAGCGCTTCCACCACCAAGGAACTGAATGCAGAGCA	19677
OY	1681	agccaagatgcagatctccaagaagaagaaagcagaagatgtgagatattgtttctacact	17400
Db	1968	AGCCAGATGCAGATTTCCAAAAAGAACCAAGACGAAAGATGAGATATTTGTTTCACCTT	20227
OY	1741	gatgcacccaaggaacttctgacatttgaagtcaccaataatgatccaagactgcagac	18000
Db	2028	GATGGCACCCAGGAATTTCTGACATTTTGAAAGTCCCACTTAATGATTCAGAGATCTCGAGCG	20877
OY	1801	cttggtgtcagtgtcacaagttacaaagttcaaaaagaaacccgcagatlttggaaatctt	18600
Db	2088	CTTGGTGTCACTGTCAAAAGGTAAACCGGTCAAAAAGAACACAGCATTTTGGGAATCTTT	21447
OY	1861	gtcaagttccattatataatgtgagagcagacatctaaagatgggaaggttctgggtgaatg	19207
Db	2148	GTCAAGTCCATTATTAATGCGAGCGACGACTCTTAAGATGGAAGGCTTGGGTGAATGAT	22077
OY	1921	caactgatagcagtaaaatgtgaagaatccctgttggcgcaagacaaaaccaagatgcattgaa	19800
Db	2208	CAACTGATAGCAGTAATAATGTGAGATCCCTGTTGGCAAGCAAAACCAAGATGCCATGGAA	22677
OY	1981	accctaagaaggtctatagtctacttgaaggaataaaacggggaatgtatccagtattgtt	20400
Db	2268	ACCCTAAGAAAGGTTATGTCTACTGAAAGCAATAAACGAGAAATGATCATCGATTATTTT	23227
OY	2041	gcaagagaaataagcaagtgcaatgagcttgaatgcacccggagacccccctgcagctag	21007
Db	2328	GCAAGGAATAAACAACAGTGCATGAGCTGAATCACTACCTGGGAGCCCCCTGTGACCTCAG	23877
OY	2101	ctgcccatgtgaacacgcttggatgtatagagaacgagaataattcccatctcccttaagt	21600
Db	2388	CTGCCCATTTGAACAGGTTGTGATGTATAGAGAACGAAAGATTTTCCATTTCCCTCTACAGT	24477

QY 2161 gggatigaagggcttgatgaatccgacgaagaatgctgcctcagtaagatagtgat 2220
 Db 2448 GGGATTTGAGGGGCTTGAATGATCCCGACGAGAAATGCTGCTCCTAGAGTAATG --- 2504
 QY 2221 gagtcagtaataaccagctgtccctacagtaataatgcccagaagttgcaactgtcatt 2280
 Db 2505 -----GTTAAATTCACAGCTGTGCTCCCTTCAGTGAATATGCCCCAAGATGACATGTCAT 2558
 QY 2281 atagaagaatgacaggtttgacagtgcttcctccacatctcctgacagctcctccaga 2240
 Db 2559 ATAGAAGATGACAGCTTGCAGTCTTCCTCCACATCTCTCAGCAGTCTCTTCCAGC 2618
 QY 2341 tcccatatgatgttgaggtttgtgagggcgagatgctgtgacttgggccaaggtgcaatc 2400
 Db 2619 TCCCATATGATGTGAGGCTTTGTGAGGGCGAGATGTGTAATTGGGCGCAAGCTTCCAGC 2678
 QY 2401 agtgcatacagcagctgtccttctgacagatgctgacagttgcttcttcaaga 2460
 Db 2679 AGTGAATTCACGCTGCTTGTGAGTCCAGATGTCAGATTTCTTCTTTCAACGA 2738
 QY 2461 gaaggaattgagctgacagtgatgtcagaanaaagcacaagaacatttcagatgcca 2520
 Db 2739 GAAGGATTTGGAGCTGAGATGTCAGAAAACGACAAACCAATTTTCAGATGCCAGT 2798
 QY 2521 caattgagattcgtlaaacaacgaanaatcaaaaagcatgtttaggtgatactgacag 2580
 Db 2799 CAATTGATTTGCTTAAACACGAAATCAAAAAGCATGATTTAGATAGTACTGACAG 2858
 QY 2581 actaaactaatacagtgatgacgaagaagaagttctccacagagatgtggctcct 2640
 Db 2859 ACTAAACTAATAACAGTGTGATGACGAGAAAGCAGGCTTCCACAGAGATGTGGCTCT 2918
 QY 2641 tccctgagctcgaagaagtcgaagctcgttggagagctgacagcagctgtgcagagtg 2700
 Db 2919 TCCCTGAGCTCTAAGAAGTCAAGCTCATTGGAGCTGTGACAGCCGAGTTCGAGGCTG 2978
 QY 2701 acttggatggagatattcctccttcacgtcgaagcgagatatacagagcgagagga 2760
 Db 2979 ACTTTGAATGGGATATTCCTTCCATGTCGACGCGCGCGATATACAGAGCGAGGGA 3038
 QY 2761 tgcatagagcttcaagagctgacatgacaaactcttataaaccgcggtatgat 2820
 Db 3039 TGCATGAGAGCTTCAAGCTGCTCCATGCACAAACTTATGATTAACCCGCGGTAGATGAT 3098
 QY 2821 gatgataagcagcatggaagacttgaagaagaacacagaagaagttaagaatacagga 2880
 Db 3099 GATGATGAAGCATGAGACCTTGGAGAAGACACAGAAAGTTCAGATCAGGGAGA 3158
 QY 2881 gactctgataccacagcagtgatcagccttccactctctgagagacaaatgaaatgga 2940
 Db 3159 GAGTCTGATCCACAGCCAGTACAGCTTCCACTCTCTGAGAGACAAATGAATGCA 3218
 QY 2941 aacaaagagaagtgataagactgataagaagaagaataaactgtaaaagaagaagaag 3000
 Db 3219 AACCAAGAGAAAGTGTATTAAGACTGATAGAAAAGATTAACCTGAAAAAGAAAGAG 3278
 QY 3001 aaagataagagataagaagaagaataaagaagaagaagaagaagaagaagaagaaga 3060
 Db 3279 AAAATATGAGATTAAGAGAGAGATAAATGAAGCCAAAGAGGAATGCTGAAGGCTTG 3338
 QY 3061 ggaagacatgttcaggtttgccaac 3085
 Db 3339 GGAGACATGTTTCCAGCTTGCCTTCCAAAC 3363

RESULT 11
AY026057

LOCUS AY026057 4167 bp mRNA linear ROD 20-FEB-2001
 DEFINITION Mus musculus PAR-3 180 kDa isoform (Par3) mRNA, complete cds.
 ACCESSION AY026057
 VERSION AY026057.1 GI:12965348
 KEYWORDS house mouse.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 4167)
 AUTHORS Lin,D., Gish,G.D., Songyang,Z., and Pawson,T.
 TITLE The carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif
 JOURNAL J. Biol. Chem. 274 (6), 3726-3733 (1999)
 MEDLINE 99121117
 PUBMED 9920925
 REFERENCE 2 (bases 1 to 4167)
 AUTHORS Lin,D., Edwards,A.S., Fawcett,J.P., Mhamali,G., Scott,J.D. and Pawson,T.
 TITLE A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and apkc signalling and cell polarity
 JOURNAL Nat. Cell Biol. 2 (8), 540-547 (2000)
 MEDLINE 20394297
 PUBMED 10934475
 REFERENCE 3 (bases 1 to 4167)
 AUTHORS Lin,D.C., Jr. and Pawson,T.J.
 TITLE Direct Submission
 JOURNAL Submitted (28-JAN-2001) Programme in Molecular Biology, Samuel Lunenfeld Research Institute, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

FEATURES

source location/Qualifiers

gene

1. 4167
 /organism="Mus musculus"
 /strain="NIH Swiss"
 /db_xref="taxon:10090"
 /gene="Par3"
 110. 4111
 /note="Par3: scaffolding protein"
 /codon_start=1
 /product="PAR-3 180 kDa isoform"
 /protein_id="AAK07669.1"
 /db_xref="GI:12965349"

CDS

translation="MKVTYCFGRPPVPCDGRMKVPSLIQDAVTRYKAVANDPNY

ITPGSLRPHGCGCITLDDILCDVADDDKRLVAFFDDPHHGGDGTASPTQSEPE
 SDEEPRKNPPTWSTTAFGLKONTAGSPKTDKRDENYRSLPDPSPMSWQPDRA
 RSLSLASHPVNDRLMEKODEDEGESESSVEVEVGHADGLEMPNPSLDDMYKLYO
 VPNDGPIGHVPEFSGARGRTGLVLYRLEKKGAKOENLFHENDCTVRLNDGDLN
 RFEDQAOHFOAMARARYTMWVYPAANKROYBLSOREKNNSPGEFSDSRVANR
 SVANNAPOALPRAPRLSOPQDLAHRPLPSHASTRPPAALPAPSVLSTNVGSV
 VNTKVKCRNLIOIKKTGEGISITSRDVITGSAPIYKNIIPRCAAIQDSRLKAG
 DRLFEVNCVDLAGKSQEEVYSLSTKMEGVSLVROEAFHPRNMAPEPOMOTP
 KETKAEDEDVLTDPGTREFLTPEVPLNDSGAGLVGVRGNRKEHNAELGFEVSI
 INGAASAKDGLRYNDOLIAVNGESLIGKAOEMETLRSMSSTEGKRMQIOLIVAR
 RISQCNELASPGSPAPELPLETDLDRRRIRISLSYSGTREGIDESTRRAALSRLMG
 KQDCLSPYNNMHPDVTMIEDDRLPVLPPLHSDOSSSSSHDVGITMEACTAKATIS
 DSADCSLSPVDVPIARORRGFGNMSERKTRFOVSASOLDVFKTKSSMDLVAD
 TKLNIYDDORAGSPSRDVGPSLGRKSSLEEDTESSRSRGSVSTSSDPSRIIRG
 RGNESFRALIDKSYDKPVYDDDEGEMETLEEDTESSRSRGSVSTSSDPSRIIRG
 OMNEDPEPRKDRTERKKDKKDKREKREKDKLAKGMLKGLGMDFRGKRRKDK
 MEKMGRIKIDSTSEEDRYRMKEORIOAKTREPERRAREROYALIODPRTGCG
 DDELLYGMSYEGCLALNAPQSPREGHMDLIYAOVKPRPSKSGDASRSRPSND
 RIORLROFOOAKODEVEDRRTYRYSFOSWSSSRPASOSGRHSYVEVVOVORER
 RESFOAOROYSLPPOSRRKNASSISDSEQONTAAVEGQSKENRRSSYGSNG
 YLGGHGNARVYMETQELLROBORRQDLKOPPADGVKRPDPDPSPSOVARLN
 RLORPEKGRPFYS

BASE COUNT 1145 a 1079 c 1174 g 769 t
 ORIGIN

Query Match. 51.9%; Score 2952.2; DB 10; Length 4167;

Best Local Similarity 83.4%; Pred. No. 0;

Matches 3442; Conservative 0; Mismatches 613; Indels 72; Gaps 5;

QY 1 atgaagtgacgtgtgtcttcgagcagaccgggtgtgtcgtgcgtgacgacggccac 60
 Db 110 ATGAAGTGAACCGTGTCTTCGGAGAGACCCGGGTGTCTGCTGCGGAGATGCGCCG 169

QY	61	atgaagaatttcagccctcatccagcgcggtgacccgctacccgaaggccatgcgaag	120
Db	170	ATGAAATTTTTCAGCCTCATCCAGCAGGGGGTGAACCCCTACCGGAAGCCGTGGCCAA	229
QY	121	gatccaactctacgtgataacaggtgctgcgtttggaaacatggagatggggaaatactaac	180
Db	230	GATCCAAACTACTGGAATACAGGTGCATCCGTTGTGACCAATGAGATGAGGATTTCTPAAC	289
QY	181	cttatacatalctcttctgtgtgtagaagacagataaagaacagactgtgtacgtttgat	240
Db	290	CTGATATGACATCCCTCTGTGACGTTGGTGATGACAAAGACAGACTGGTAGCATATTTCAT	349
QY	241	gagcaggaaccacaatcaacgaggtgtatgtgacccagtgtccagttccacggttacccagagc	300
Db	350	GAMCAGGATCCCCCATGTGAGGAGATGTGACAGCGCCAGCTTACCGGAACCCACAGAT	409
QY	301	ccgaagaatttgtgtagtgtgcttggcaccaacatgtctacgcttccagcttcaagcttaca	360
Db	410	CCAGAGATTTCCGCACTGTAGCTGGGACCAACAAATGTTTCGTCTTTCAGCTTATCAA	469
QY	361	gcaacaaagttaaaattgagtgacacacctcaagtccttccttcgacgaatatgcttccatgt	420
Db	470	GCCACAAAGTGAATTTGAGGTGCACGCGCTTGAGTTCTTCCGGGAAATATGCTTTATGTGC	529
QY	421	cgaagcagtgatgtaccacgctctaattgtgcctctccacttcgtcaagtgtaatctt	480
Db	530	CGCGGAGCAGCGACCCACAGTTTAACTGGCGTTTCCACTTCTGTAGTATMACACTTT	589
QY	481	tcctcttaagaagccttcaagaaataatccacaagctgtgtcaacaacagctgtcctctc	540
Db	590	TCCTCAAGAGACCCCTTCAGGAAAAACCCACCCGCTGTTCACACGACACTGGCTTTTCTC	649
QY	541	aagcagaacactgtctggagttcctaaacactgacagaagaaagaaatgtaaaaactacaga	600
Db	650	AAGCAGAACACCGCTGGAATCTCCAAACCTGCGACAGAAAGATGAATAACTACAGA	709
QY	601	agctcccgagggtactagttaactgtgtctaaaccaattccaagagacaaatgtcgtccg	660
Db	710	AGCCTTCACAGGGATCCCACTACTGTGTCACCACTTCACAGAGAACACCCCGCTCC	769
QY	661	tctctgagttccagtcaccccaactgtgtgggcaagtgtgctgtgaagaacaaagaacagatgag	720
Db	770	TCCTTGAGGCCAGCCACCAATGTATGAAACCGGTGCTGTGAGAGCAACAAACAGATGAG	829
QY	721	gatgtgacagaagaagataacagtcgtgttgaacctgtttgacatgtgtacacgggtttg	780
Db	830	GAAAGCAGACAGAAAGACACACACCCAGTGAGACCCGTTGGACATGCTGATACCGGATTTG	889
QY	781	gagcatatacccaacttctctctgtgtatgtatgtaaagctgtgtgaagttcccaagat	840
Db	890	GAGAACTATGCCACTTTTCCCTCGATGATGTATGTAAAGCTGTATCAATCCCAACGAT	949
QY	841	ggaaggcctctggyaaltcca tgtagtcccttccatgaigtgcagggcgycgacgaacctgggg	900
Db	950	GGAGGGCCCCGGGAATTCATGTATGTCCTTTCAGTGTGCGAGGGCGGCAAACTTTGGGG	1009
QY	901	ttattagtaaaaagatattggagaaagttgtataaagctgtgaacatgaaaatcttlttcgtgag	960
Db	1010	TTTGTACTGAAGCGGTTGGGAGAAAGCGGTGAAGGCTGACAGCAAGAAAACTTTTTCATGAG	1069
QY	961	aatgattgcattgtcagagattaaigtatgtgagaccttcggaatagaagaattgttaaaga	1020
Db	1070	AATGACTGTGATGTGGAGATTTAAAGATGAGATCTTCGAAACAGAAAGATTTTAGAGACGA	1129
QY	1021	caacatatgtttcgccaagccatgctgtacacccatcatittgttccatgtgttctgtca	1080
Db	1130	CAACATATGTCTCCGCAACGTATGCGTGCAGCTGTCAATTTGGTTCACTGTGTCCTGTGA	1189
QY	1081	gcaataaaggcgagtatgaacaactatcccaaatgtgagaagaacatatctactatcaagc	1140
Db	1190	GCAAAACAAGGAGCAATATGAAACAACATGTCCCAACGCGGAAGAACAACATATCTCCAGGCG	1249

QY	1141	cgcttgccttcgacagcagtatattgcacaaaggaagfqtgaacgfygcagygctttac	1200
QY	1141	cgcttgccttcgacagcagtatattgcacaaaggaagfqtgaacgfygcagygctttac	1200
Db	1230	cgcttgccttcgacagcagtatattgcacaaaggaagfqtgaacgfygcagygctttac	1309
QY	1201	acgfygcagagagaccgcagctagtaacaccacccgcctgcagacagatagacttcaactcaaga	1260
Db	1310	gcatttgccacagaccacccagctgagtcacaccaccggacacacttgatggtctacaccccca	1369
QY	1261	ctacctatagcgacacacccctcgagaaacacacalcgcgtccagcttcggacactcag	1320
Db	1370	ctacctatagcgacacacccctcgagaaacacacalcgcgtccagcttcggacactcag	1429
QY	1321	aatgattattagtaagctgtcaagcagfygttataacacaaaanaatgagcaagagctt	1380
Db	1430	agtgctgcttaattaccacacgtraggcagctgtrctacacacacaaagtaagcagagcttc	1489
QY	1381	aatalccagctttaaagaaaggtacagaagfcttggatcttaagcatcacttccagagatga	1440
Db	1490	aacatccagctttaaagaaaggtacagaagfcttggatcttaagcatcacttccagagatga	1549
QY	1441	acaaatagtggtctagcttccaactctatgtgaaanaaatcttcccccgaggcgccactt	1500
Db	1550	acaaatagtggtctagcttccaactctatgtgaaanaaatcttcccccgaggcgccactt	1609
QY	1501	cagagtgccgcacttaagcagagagacagacactatagtagtlaatfygagfagattttag	1560
Db	1610	cagagtgccgcacttaagcagagagacagacactatagtagtlaatfygagfagattttag	1669
QY	1561	gagcaaatcccaagaagaagatgttctgcgtgtgtagaagcaccaagaatggaagaactgtg	1620
Db	1670	gagcaaatcccaagaagaagaagatgttctgcgtgtgtagaagcaccaagaatggaagaactgtg	1729
QY	1621	agcccttcgyccttcgcagagaaagagcccttcaccacaaaggaactaaatgcagaagca	1680
Db	1730	agcccttcgyccttcgcagagaaagagcccttcaccacaaaggaactaaatgcagaagca	1789
QY	1681	agccagatgcagatctccaaaagaacgaacagcaagatgagagatattgttcttaacact	1740
Db	1790	agccagatgcagatctccaaaagaacgaacagcaagatgagagatattgttcttaacact	1849
QY	1741	gatggcaccaggaattcttcgacatttgaagctccacttaatgatcttcagagatcttcgagc	1800
Db	1850	gatggcaccaggaattcttcgacatttgaagctccacttaatgatcttcagagatcttcgagc	1909
QY	1801	cttgytgcagtgtcacaaaggtlaaacccggttcacaaagagacacacgcagatttggaaactctt	1860
Db	1910	cttgytgcagtgtcacaaaggtlaaacccggttcacaaagagacacacgcagatttggaaactctt	1969
QY	1861	gtcaagttccatattatattggaagagcagcatctcaagaatggaaagcttcgggtgaatgat	1920
Db	1970	gtcaagttccatattatattggaagagcagcatctcaagaatggaaagcttcgggtgaatgat	2029
QY	1921	caactgatagcagttaaatgtagaatccctcttggcagaagacaaacaaagatcccatgtaa	1980
Db	2030	caactgatagcagttaaatgtagaatccctcttggcagaagacaaacaaagatcccatgtaa	2089
QY	1981	accctaagaagfcttatgtctacttgtaagcaataaacsagaatgatccaagattatgtt	2040
Db	2090	accctaagaagfcttatgtctacttgtaagcaataaacsagaatgatccaagattatgtt	2149
QY	2041	gaaagagaataaagcaggttcaatgagctgtaaaatcactcggagagcccccctgagacttag	2100
Db	2150	gaaagagaataaagcaggttcaatgagctgtaaaatcactcggagagcccccctgagacttag	2209
QY	2101	ctgcccatttgaacaaagcgttggatgatataagaaacgaagaatttccattccctctacagt	2160
Db	2210	ctgcccatttgaacaaagcgttggatgatataagaaacgaagaatttccattccctctacagt	2269
QY	2161	gggatgtgggggcttgatgaatgcgccagcagaagaatgctgcgcctcaagtatagaataatggt	2220
Db	2270	gggatgtgggggcttgatgaatgcgccagcagaagaatgctgcgcctcaagtatagaataatggt	2327
QY	2221	gagtaagtaaatccagcgtctccctctacagfataatctcccaagaataaaccttcat	2280

```

Db 2328 -----GTAATGCCAGCTGTCCACACGTCACATGCTCAGCATGACACTGTCAG 2380
QY 2281 atagaatgacagattgctcagctgtctccacatctctcagcaagctctccacg 2340
Db 2381 ATTGAAGATGACAGGCTGCTGCTGCTCCTCCTCCTCCTGACAGCTCCTCCTCCTC 2440
QY 2341 tcccatgatagtatggtggttctgtacgagagatgctgtaactggtgcaagctgcaac 2400
Db 2441 TCCCATGATGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2500
QY 2401 atgatacagcagcagctgcttctgagctcagatggttgcagcttcttcaaga 2460
Db 2501 AGGACTCAGCCGACTGCTCCTGATGATGATGATGATGATGATGATGATGATGATG 2560
QY 2461 gaagagattgagactcagagatgctcagagaaacagcaagcaatttccatccagt 2520
Db 2561 GAAGGATTTGAGAGCGAGATGATGATGATGATGATGATGATGATGATGATGATG 2620
QY 2521 caattgatttctgtaaaacacgaaatcaaaagcatgatttagtataagctgag 2580
Db 2621 CAATTGATTTCTGTTAAACACGAAATCAAAAGCATGATTTAG---TACTGACGAG 2677
QY 2581 actaaactcaatagatgatacgaagaagagatttccagagagagatggtgtcct 2640
Db 2678 ACTAAACTCAATACAGTGTGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2737
QY 2641 tccctgagctcgaagaagctcagctgtgagagctcagagagagctgagagtg 2700
Db 2738 TCCCTGAGCTGCTGAAAGAAATCTGATCTGATGATGATGATGATGATGATGATGATG 2797
QY 2701 acttgaatgagagatcttccatctctcagctcagcagcagcagatgataagagagag 2760
Db 2798 AGCTGGAACGGGAAATCTCTCTCTCCAGCCCTCGGCGCAAGATATCTCGGGGAAGGG 2857
QY 2761 tgcagtagagctcagagctcagcagctcagcaatctctatgataaaccgctgagatg 2820
Db 2858 TGCAACGAAAGCTTCAGAGCGCGCATGACAAAGTCTGACAAAGCCATGCTGATGAT 2917
QY 2821 gatatagagagcagagagacttggagaagacacagaagaagtlcaagatcagagaga 2880
Db 2918 GACGACGAGGCTATGAGACCTTGTGAAGAACACAGAGAGAGAGAGAGAGAGAGAGAG 2977
QY 2881 gactgtatcacacagcagctcagcagctcctcagctcctgagagagagagagagag 2940
Db 2978 GAGTCCGCTGCTCAGCTCAGGATCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3037
QY 2941 aacaaagagagagctgataagactgataaagaagataaactggaagaagaagaag 3000
Db 3038 GACCCAGAGAAAGGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3097
QY 3001 aaagataagataagagagagagagagagagagagagagagagagagagagagag 3060
Db 3098 AAAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3157
QY 3061 ggaagacatgctcagcttctgcaaacatcgaagaagatgagaagagagagagagag 3120
Db 3158 GGGGACATGTTCAAGGTTTGGCAAAACATGCAAAAGATGACAAAGATGAAAAAGGTTG 3217
QY 3121 ataaaaataagagatctctcagcagaagaagagagagagagagagagagagagag 3180
Db 3218 ATAAAAATTCAGATTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 3277
QY 3181 gagaagattcaagcacaactcgaagaatttaggagagagagagagagagagagagag 3240
Db 3278 GAGAGGATTTAAACCAAACTCCAGAGTTTAAAGGAGCGGCAACCGCAGAGCTGACTAT 3337
QY 3241 gctgaatcaatcaatctcagcagatctgctgagatgagatgagatgagagagag 3300
Db 3338 GCAGAGATTCAGAGATTTCAATCGAGCTTGGCTGTGATGACAGAGTTGGCTGTGG 3397
QY 3301 gttctctatataagagctcagctcagagagagagagagagagagagagagagagag 3360

```

```

Db 3398 ATGTATCTCTATGAAGGCTGCTTGGCTCTCAATGSCCAAGACCCAGAGCCCAAGAGAGG 3457
QY 3361 catalgataagatgcttctgatalgcccagtlcaagaagccgagagatctccaaaccctcact 3420
Db 3458 CACCTGATGAGACACTTGTGATGCTCCCAAGTAAGAAACCTCGAGAGTCCAA-----ACCT 3511
QY 3421 gtagagagtaagatcaactctcagcaatcagatgagatgagatgagatgagatgag 3480
Db 3512 GGAGACACCAATTCAGTCACTCTGCAACCATGACCGGATACCGCTGACGCAAGAG 3571
QY 3481 ttccagcaagcaagcaagatgataagatgataagatgataagatgataagatgataag 3540
Db 3572 TTCCAGCAAGCAACCAAGATGAGATGATGATGATGATGATGATGATGATGATGATG 3631
QY 3541 caaccctgagcagagcagagcagagcagagcagagcagagcagagcagagcagagcag 3600
Db 3632 CAATCTTGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 3691
QY 3601 gtgagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3660
Db 3692 GTTCAAGTACAAAGGAGCGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3751
QY 3661 taccgctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3720
Db 3752 TACAGCTCAGCTCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3811
QY 3721 gagaagagatctcctcctgagagagcctcagagagagagagagagagagagagagag 3780
Db 3812 GAACAGAGATCTAGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3871
QY 3781 agtacaagagctcagagagagagagagagagagagagagagagagagagagagagag 3840
Db 3872 AGTTACAGAGGCTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATG 3931
QY 3841 ctggaactcaagagagcctcctcagcagagagagagagagagagagagagagagagag 3900
Db 3932 CTGAGAGCCAGAGAGCTTCTCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3991
QY 3901 caagctcctcagagagagagagagagagagagagagagagagagagagagagagag 3960
Db 3992 CAGCTCAGCTCAGTATGAGTACAG----- 4015
QY 3961 taagcctcctcagagagagagagagagagagagagagagagagagagagagagag 4020
Db 4016 -----GGGCCCTTCGCGCAAGATGATGATGATGATGATGATGATGATGATG 4060
QY 4021 aggttaagagagatcagactcctgagaagaagagagagagagagagagagagagag 4080
Db 4061 AGCTGAAACAGACTGACAGACCGGAGAAAGGCGGCGCTTCTACCTGAGAGCTCAAGA 4120
QY 4081 taacag---atgctcatgtcgcgcaataaagaacatlttctatga 4124
Db 4121 AGGATGGCAGCATTTTGTGTGATGCAATTAAGGACATTTATCCCATGA 4167

RESULT 12
AF332592 2967 bp mRNA linear PRI 16-NOV-2001
LOCUS AF332592
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein
ACCESSION AF332592
VERSION AF332592.1 GI:14579308
KEYWORDS Down-regulated expression of atypical PKC-binding domain deleted
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2967)
AUTHORS Fang,C.M. and Xu,Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE 21520080

```

PUBMED 11642408
REFERENCE 2 (bases 1 to 2967)
AUTHORS Fang, C. and Xu, Y.
TITLE Exon/Intron Structure And Splicing Variants of A Novel Human Polarity Gene, hasip
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2967)
AUTHORS Fang, C. and Xu, Y.
TITLE Direct Submission
JOURNAL Submitted (29-Dec-2000) Laboratory of Molecular and Cellular Oncology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P. R. China

FEATURES
source 1. .2967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1. .2967

CDS

/note="ASIPSB; alternatively spliced; similar to GenBank Accession Number AF196186"
/product="atypical PKC isootype-specific interacting protein short variant b"
/protein_id="AAK69192.1"
/db_xref="GI:14579309"
translation="MKTVVCFGRTRVYVPCGDGHHKVPFLIQAVTRYRKAIARQPNY
WIOVRLHKGSGILIDLDILCDVADDDRIYVREDODPHGGDGTASSGTCQSP
IFGSELGNNVSAFOYQATSEIETPVLNANPLHRRSSDPLILSTSVSDNF
SEEPKRNPTKMSSTAGLQKNTAGSKTCDKRDENYRSLPRTSMNOFORDNA
RSSLSASHPWKNMLEKQEDQEDGETEDNSVEPEVHADTGLEHLPNSLDMVLE
VPNGGPGLGIVHPFSAAGKRTGLIVRLKEKGSAEHEHNLREMDCLVINDCLRN
RRFOAOHMFROAMRTPIIFPVVPAANKEDYEOLESEKNVYSRSPSOYIDNR
SVNSAGLHTVORARLNHPROIDSHSLPSAHPSGKPPAPASAPQNVSTVSSG
YNTKIKKRLNIQAKKEGEGFSTSDVTIGSAPITYVNIILPRGAIIDGRKAK
DRLIEKGVLDVKSQSEVSVLSLSTKEGVSLVLFQEDAFHRELKADBEDVLT
PDGREFLTFEPLNDSGAGLVSKNREKNENADIGIVKSLINGANSKDLRL
VNDLIANGESLIGKTQMDAMETLRMSTEGNKADIGIIVARIKCNELKPGS
PGPELPLETALDRERISHSIVSGIGLDESPRNALSRIMKCYOLSPVNNPD
DTVIEDRLPLVPLPHLSDSSSHDVGVTADGAWAAAIISDADCSLSPVVD
VIAFORBEGFQIADERTKLTNYVDOKASPSRDVPSGLKSSLESLEQTVAEVTL
MGDIPEHPRPRTIRGRCNESFRALIDKSTDKPAVDDDEMELEEDTESSSGR
ESVSTASDQPSHSLERDMNGNOEKDKTDRKKDKKRRKKDRKKRKKRKKMLK
GLDMFSLAKLKPERK"

BASE COUNT 890 a 653 c 776 g 648 t
ORIGIN

Query Match 48.7%; Score 2772.6; DB 9; Length 2967;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 2943; Conservative 0; Mismatches 4; Indels 138; Gaps 3;

QY 1 atgaagatgacccgtgtgcttcggaagcagcggtggtgctgacgtgaggaagcagc 60
DB 1 ATGAAAGTACCGTGTCTCGAGGAGCCCGGGTGTCTCGCGGAGACGGCCAC 60
QY 61 atgaagtttcacgctcatccagcaggcgtgacccgctaccggaagcctacgcgaag 120
DB 61 ATGAAAGTTTACGCTCATCCAGCAGCGGTGACCCGCTACCGGAGGCCATTCGCCAAG 120
QY 121 gatccaactactgatacaggtgcatcgtctggaacatgagatgagaaactactagac 180
DB 121 GATCCAACACTGATACAGGTGATGATGCTTGGAAACATGGAGATGGAGATACAGAC 180
QY 181 cttagatgacatctcttggatgtagcagaagataaagacagcgttagcagtggttgc 240
DB 181 CTGATGACATCTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 gagcagatccacatcagcagagtgatgacacagtgccaggtccacagcagcagc 300
DB 241 GAGCAGATCCACATCAGCAGAGGTATGGACACATGTCAGTTCCAGCGGTACCCAGAC 300

QY 301 ccagagatatltgtagtagcctggaaccaacaatgctcagccttcagccttaccaa 360
DB 301 CCAGAGATATTTGGTAGTAGCTTGACACCAAAATGTCCAGCCTTTCAGCCTTACCA 360
QY 361 gcaacaatgtaaatgtatgacacactcagtccttcagcgaataatgtccttcatt 420
DB 361 GCAACAATGTAATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 cgaacagtagtagcagcagcctcattgagccttcagccttcagccttcagccttc 480
DB 421 CGACGAGTAGTAGCAGCAGCCTTAAATTTGAGCCTTTCAGTGTATGATTTT 480
QY 481 tctctgaaagagccttcaggaataatccacacagcgtgtaacacacagcgtgcttc 540
DB 481 TCCTGTGAAGAGCCTTCAAGGAAATCCACACGCTGTGCAACACAGCTGCTTCCTC 540
QY 541 aagcagaacactgctgaggttcttaaaactgtgacaggaagaagaatgaaactaca 600
DB 541 AACGAGAACACTGCTGGAGTCTTAAACCTGGACAGGAAGAAATGAAACTACAGA 600
QY 601 agcctccgagagatactagtaactgtgtcctaaccaatttcagagagacaatgctgc 660
DB 601 AGCCTCCGCGGATTACTAGTAACCTGTCTAACCAATTTAGAGACAAATGCTGCTGC 660
QY 661 tctctgagtagcagtcacccaatggtggaagtggtcgtgagaacaagaacagatgag 720
DB 661 TCTCTGAGTAGCAGTCAACCAATGTGTGGCAAGTGGCTGAGAAACAAAGAGATGAG 720
QY 721 gatggagcagaagaagataaagtcgtgtgtaacactgttgacaatgctgacaggttg 780
DB 721 GATGGAGCAGGAAGATTAACATGCTGTGAACCTGTGGACATGCTGACAGGGTTG 780
QY 781 gaacataaccacacttctcgtgataatgtaagctctgagaagcctcccaacgat 840
DB 781 GACATATACCAACTTTCTCTGATGATATGTAAGCTCTGTAAGCTCCCAACGAT 840
QY 841 ggaaggcctctggaatccatgtagtgccttcagtagtcgaagcagcgaacacctgag 900
DB 841 GGAGGCGCTCTGGAAATCCATGTAGTCCCTTTCAGTCTGAGCGGACAGACCTGGGG 900
QY 901 tctttagtaaaagatttggaagaagtgtgaaagctggaataagaatactcttcgtgag 960
DB 901 TCTTGTAAAGGATTTGGAAGAGTGTGAAGCTGMAATGAAATCTTTTCGGAG 960
QY 961 aatgattgcatctgcaagatgaaatgtagtgcaccccttcgaataagaatttgacaaga 1020
DB 961 AATGATTGCAATTTGCAAGATTAATGATGCGGACCTTGAATGAAATTTGAACAAACA 1020
QY 1021 caacataatgttcgcaagcactgctacacccaatcattgctcattgctcattgca 1080
DB 1021 CAACATATGTTTCCCAAGCAGCATGCGTACACCCATCATTTGTTCCATGTGTTCCGCA 1080
QY 1081 gcaataaagagagatgataaacaactatcccaagtgaagaagaacaaatctatcaagc 1140
DB 1081 GCAAAATAAAGAGATGATGAACAACTATCCCAAGGAGAAACAATTTCTATTCAAGC 1140
QY 1141 cgttttagcccttcagcagcagatlattgacaacaggagtgtgaacagttgagagcttcac 1200
DB 1141 CGTTTACCCCTCAGACCGCATATATTGACACAGAGTGTGAACAGTGCAGGCTTCAC 1200
QY 1201 acggtgcaagagacacccgactgaacacccgctgtagcagataagacttcactcaaga 1260
DB 1201 ACGGTGCAAGAGACACCCGAGTGAACCAACCCCGTAGAGATATACCTCTCAAGCCTCAAG 1260
QY 1261 ctaactcatagcagacacccctcgggaaacaacacatccgcttcagcctcagcctcag 1320
DB 1261 CTACCTCATAGGCAACACCCCTCGGAAAAACCACTTCGCTCCACCTCGGCACTCAG 1320
QY 1321 aatgataatgtagcagctgtaagcagtggttataacaacaaaataagcagaagctt 1380
DB 1321 AATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 aatatacagcttaagaagaagttaggaattcagcatcacttcagaagatgta 1440

Db 1381 AATATCCACCTTAAGAAAGGTACAGAGGTTGGATTGACATCCTCCAGAGATGTA 1440
 Oy 1441 ACAATAGTGTGCTAGCTCCAAATCTATGTGAAAAACATTCCTCCCGGGGGGCCATT 1500
 Db 1441 ACAATAGTGTGCTAGCTCCAAATCTATGTGAAAAACATTCCTCCCGGGGGGCCATT 1500
 Oy 1501 CAGGATGCGCACTTAAGGACAGAGACTTATAGAGTAAATGAGATTAAGT 1560
 Db 1501 CAGGATGCGCACTTAAGGACAGAGACTTATAGAGTAAATGAGATTAAGT 1560
 Oy 1561 GGCATATCCCAAGAGAGAGTTGTTGCTGTGAGAGACCAAGATGAGAGACTGTG 1620
 Db 1561 GGCATATCCCAAGAGAGAGTTGTTGCTGTGAGAGACCAAGATGAGAGACTGTG 1620
 Oy 1621 AGCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1680
 Db 1621 AGCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1680
 Oy 1681 AGCCAGATGCAATCCCAAGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATG 1740
 Db 1681 AGCCAGATGCAATCCCAAGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATG 1740
 Oy 1741 GATGAGACAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1800
 Db 1702 GATGAGACAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1800
 Oy 1801 CTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1860
 Db 1762 CTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1860
 Oy 1861 GTCAAGTCCATTAATTAAGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1920
 Db 1822 GTCAAGTCCATTAATTAAGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1920
 Oy 1921 CAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1980
 Db 1882 CAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1980
 Oy 1981 ACCCTAAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2040
 Db 1942 ACCCTAAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2040
 Oy 2041 GCAAGGAGATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2100
 Db 2002 GCAAGGAGATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2100
 Oy 2101 CTGCCATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2160
 Db 2062 CTGCCATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2160
 Oy 2161 GGGATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2220
 Db 2122 GGGATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2220
 Oy 2221 GAGTCAAGTAAATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2280
 Db 2179 GAGTCAAGTAAATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2280
 Oy 2281 ATAGAAGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2340
 Db 2233 ATAGAAGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2340
 Oy 2341 TCCCATGATGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2400
 Db 2293 TCCCATGATGATGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2400
 Oy 2401 AGTATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2460
 Db 2353 AGTATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2460
 Oy 2461 GAGATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2520
 Db 2412 GAGATTAAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2520

Db 2413 GAAGATTTGGACGTACAG----- 2430
 Oy 2521 CAATTGATTTCTGTTAAACACGAAATCAAAAGCATGATTAGTATAGCTAGAG 2580
 Db 2431 ----- 2580
 Oy 2581 ACTAATCAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2640
 Db 2443 ACTAATCAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2640
 Oy 2641 TCCCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700
 Db 2503 TCCCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700
 Oy 2701 ACTTGTGATG 2760
 Db 2563 ACTTGTGATG 2760
 Oy 2761 TCAATG 2820
 Db 2623 TCAATG 2820
 Oy 2821 GATG 2880
 Db 2683 GATG 2880
 Oy 2881 GATG 2940
 Db 2743 GATG 2940
 Oy 2941 AAGATG 3000
 Db 2803 AAGATG 3000
 Oy 3001 AAGATG 3060
 Db 2863 AAGATG 3060
 Oy 3061 GAGATG 3120
 Db 2923 GAGATG 3120

RESULT 13
 LOCUS AK000761 2696 bp mRNA linear PRI 22-FEB-2000
 DEFINITION Homo sapiens CDNA FLJ20754 f1s, clone HEP02246.
 ACCESSION AK000761
 VERSION AK000761.1 GI:7021050
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens hepatoma cell_line: HepG2 CDNA to mRNA, clone lib: HEP
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,
 Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,
 Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
 Nakamura, Y., Isogai, T. and Sugano, S.
 NEDO human cDNA sequencing project.
 JOURNAL Unpublished (2000)
 TITLE 2 (bases 1 to 2696)
 REFERENCE Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
 Shibahara, T., Tanaka, T. and Nakamura, Y.
 AUTHORS Direct Submission
 JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
 Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@ims.u-tokyo.ac.jp,
 Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; CDNA full insert
 COMMENT Sequencing: Research Association for Biotechnology; CDNA library


```

|||||
Db 1682 TGGTACACATGGGGGTCCTTTTGAACGCTTATGACTCAACATTAACATTAAGTAT 1741
|||
Qy 4725 gatcctgtcgcgtttataataacgcttctgtcgtctgtctgaacgtgggttag 4784
|||
Db 1742 GATCCTGTGCTGGCTTATATATAGCTTGTTCGCTCTGATCCGACACGTGGTGTAG 1801
|||
Qy 4785 tcacacagcctcgtgtgtgggaacgtgggaacagatgctcctcgcgggggaagctg 4844
|||
Db 1802 TCACCAACAGCTGTGTGTGGAGACGTGGAGACAGAGATGGCTCTCGCCGGGGAGCTG 1861
|||
Qy 4845 ggcctgcaatgtgcctgtgtctatcatagggggagagctaagaagaattctcctagg 4904
|||
Db 1862 GGCCTGCATTTGGCCCTGCTCTATCATAGAGGGGAGAGCTAAGAAAGAAATTCCTAGG 1921
|||
Qy 4905 aagagctatagcccgagacactcctcctcctgaattatattatagttttgtctcgaact 4964
|||
Db 1922 AAGAGCTCATGGCCGACATCCATGTAATTAATTAATTAATTAATTAATTAATTAATTA 1981
|||
Qy 4965 tgcagagagagcagacagatgtggacagagataaacacagacagctcattgtctctc 5024
|||
Db 1982 TGTACAGGAAGGACAGAAATGGAGATGAATTAACACAGACAGTCAATTTGATCTGCTCTC 2041
|||
Qy 5025 taagggttttcaagtcaaggcaattgatgctgtctctatgcatccacacactgcatgtc 5084
|||
Db 2042 TACGCTTTTTCAGTCACAGAGCAATTTGATGCTTGTCTATGATCATCCACACACTGCATGTC 2101
|||
Qy 5085 tgcagtgagatgcacagctcctcctcctcctcctcctcctcctcctcctcctcctcctc 5144
|||
Db 2102 TGACTGGGAGATGCCACGCTCTTAAGTACTTCTGCCATTAACATTAAGACAAAGAAAA 2161
|||
Qy 5145 gccgttacacacacacagagaaacatttcggtgccacagcgtgtgtggcaggaagctc 5204
|||
Db 2162 GCCGTTACATCAATCAACACAGAGAACTTTTCGGGTGCCACAGGGGAGGAGAGAGCTC 2221
|||
Qy 5205 acctctgcgtcaagtattagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5264
|||
Db 2222 ACTCTGGGCTAGTATGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2281
|||
Qy 5285 tccctcatgttctgaaacatccctgattgtlaaacatgctgtgggtgtcctaaagtctgt 5324
|||
Db 2282 TCTCTTCAATTTCTCAACATCTCTATGATGTAAACATGGCTGGGTGCTAAAGTCCCTGT 2341
|||
Qy 5325 gaatcccgatgtgaaaaaagctgtgaggtgaaagctcagacatacctatgattacttaaa 5384
|||
Db 2342 GAATCCGATGTGAAAAAGCTGAGGTGAAAGCTCACATATCATGATTTACTTTAAA 2401
|||
Qy 5385 aacgagaaaaaagacatgatagtatgtctatatttttttttttttttttttttttttttt 5444
|||
Db 2402 AACAGAAAAAAGACATGTATGATATGCTATTTTTTTTTTTTTTTTATTTTATTTTATTTT 2461
|||
Qy 5445 tgtgttgactgttlltgaatgatgtgtcacaacacagctacccgtgtctctctcgtcalt 5504
|||
Db 2462 TGTGTGACTTGTTTTGAATAATGATGTGTCCACACACAGTACCGTGTCTCTTGCAANT 2521
|||
Qy 5505 tctgtgtcaatgtctgtcttcttaacacgtgtggcggtgtcctaagtgtgttaccagtg 5564
|||
Db 2522 TCTGTGTATGTCTGTCTTTTCTTAATCACTGTCGCGGCTGTCTAAGTGTGTGTATACAGTG 2581
|||
Qy 5565 taagcagagatgacattgtatgaacagtgctctctcctcagacgtccctcgaagctgtgaga 5624
|||
Db 2582 TACGGGCAAGTACCTTGTGATGACAGTGTGCTCTTCTCACAGGCTCCCTGACGCTGTGACA 2641
|||
Qy 5625 aacaccttctcgtacatagcaactcctaataaagaacatattctcctgt 5679
|||
Db 2642 AACACCTTCTCTGTATATGCACTCCTAATAAAGCAATATTTCTTCCCTGTT 2696
|||

```

RESULT 14
 BD004905 2718 bp DNA linear PAT 31-JAN-2002
 LOCUS
 DEFINITION Novel genes encoding protein kinase or protein phosphatase.
 ACCESSION BD004905

```

VERSION BD004905.1 GI:18632866
KEYWORDS JP 03074935-T/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
AUTHORS I (bases 1 to 2718)
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K., Otsuki,T.,
Funahashi,S., Seno,C. and Nezu,J.
TITLE Novel genes encoding protein kinase or protein phosphatase
JOURNAL Patent: JP 03074935-T 2 30-JAN-2001,
HELIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA,
KOJI HAYASHI, KAOBU SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, OMORYASU
SUGIYAMA, AI WAKAMATSU, KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI
FUNAHASHI, HIROKI SENO, JUNICHI NEZU
COMMENT OS Homo sapiens (human)
PN JP 03074935-T/2
PD 30-JAN-2001
PF 28-JUL-2000 JP 2000005061
PR 29-JUL-1999 JP 99P 248036, 11-JAN-2000 JP 00P 118776 PR
02-MAY-2000 JP 00P 183767, 09-JUN-2000 JP 00P 241899 PR
18-OCT-1999 US 60/159590, 17-FEB-2000 US 60/183322 PI TOSHIO
OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, PI KAOBU SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI FUNAHASHI, CHIHIKI SENO,
PI JUNICHI NEZU
PC C12N15/12, C12N9/12, C12N9/16, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10, C07K16/40,
PC C12Q1/68, C12Q1/42, C12Q1/48
CC
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..2718
/db_xref="taxon:9606"
BASE COUNT 840 a 586 c 688 g 604 t
ORIGIN
Query Match 43.5%; Score 2476.8; DB 6; Length 2718;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 2603; Conservative 0; Mismatches 7; Indels 102; Gaps 2;
Qy 374 ttgaagtcacaccttcagtccttcgagaacaataatgctctcattatgttcgacgagtagtg 433
Db 1 ttgaggtcacacacttcagtccttcgagaacaataatgctctcattatgttcgacgagtagtg 60
Qy 434 acccagctcctaattgtgctctccacttcctcgtcagtgatgataatttccctcctgaagac 493
Db 61 ACCCGCTCTTAATTTGGCTTCCACTTCTGTCAGTATGATTAATTTTCTCTGTAAGAGC 120
Qy 494 cttaagaaaaaatccacacacgctgtgtcaacaacagcgtggtcctcctcagaagacactg 553
Db 121 CTTCAAGAAAAAATCCACACGCTGTGTCACACACAGCTGTGCTCTCTCAAGCAACACTG 180
Qy 554 ctggagatccttaaaaccttgcagcaggaagaagatgaaaaatacagaaagcctcccgcggg 613
Db 181 CTGGGAGTCTTAACCGCTGCGACAGAGAAAGATGAAGAAACTACAGAAAGCCCTCCGGGG 240
Qy 614 ataactagtaactgtgtcctaaccattcagagagacaatgtcctcgtctcgtctcgtgtgca 673
Db 241 ATACTAGTAACCTGGTCTTACCAATTTTCAGAGAAAGAAATGCTGTCTGTGAGTGGCA 300
Qy 674 gtcacccaatgttgagggaagtgctgcygagaaacaagaagaatgaagatggagacagaga 733
Db 301 GTACCCCAATGTGTGGGCAAGTGTGCGAGAAACAAGAACAGATGAAGATGGAGGAGAG 360
Qy 734 aggaataaagtcgttgttaacctgttgacatgctgcgaacggtgttgagacatatacca 793
Db 361 AGGATTAACAGTGTGTGTAACCTGTGTGACATGTGACACAGCGGTTTGGAGCATATACCA 420

```

QY 794 accttcctcgtgataatgtaagctcgttagaagtcctcccaaga tggaggcctcgtg 853
| | | | |
Db 421 ACTTCTCTGATGATGTGTAAGCTGTGAGAGTCCCAACGATGAGGGCCTTG 480
QY 854 gaatccatgtagtgccttcagtcgtcagagcgagaaacctgggtattatagtaaac 913
| | | | |
Db 481 GAATCCAGTATGATGCTTCAGTCTCGAGGGCGCAAACTGGGTTATTAGTAAGAAC 540
QY 914 gattggagaaggtgtagaaagctgagaca tgaaaatcttttcgttagaagatgtagt 973
| | | | |
Db 541 GATTGGAAAGAGTGTGTAAGCTGAAACATGAAATCTTTTGTGATGATGATTCATTTG 600
QY 974 tcaagataatgtagtgagacctcgaaatagaaga ttcgaacaagcacacaatagtcttc 1033
| | | | |
Db 601 TCAGATTAAATGATGCGCACTTCGAAATAGAAATTTGAACAAACCAACATATAGTTTC 660
QY 1034 gccaaagcatcgtatcacccatcatttggttcacatgtgttccttcacaaataaagagc 1093
| | | | |
Db 661 GCCAAGCCATGCTGATACCCCATCATTTGGTTCATGTGGTCTCCAGCAAAATAAGAGC 720
QY 1094 agtatgaacaactatcccaagtgagaagaacaattactatcacaagccgttttagccctg 1153
| | | | |
Db 721 AGTAGAACAACATATCCCAAGTGAAGAAACAAATTACTATTCAAGCCGTTTACCCCTG 780
QY 1154 acagccagtatata tgaacaacagagagtgtgaaacagtgcagggtcacaacgltgcagagag 1213
| | | | |
Db 781 ACAGCCACTATATTCACAAACAGAGACTGTGAACAGTGCAGGGCTTCACACGGGTGCAGAGAG 840
QY 1214 caccgccatgaacaaccaccgcctgagcagatagactctcaactcaagaactactcaatagag 1273
| | | | |
Db 841 CACCCCGACTGAACACCCCGCTGAGCAATGATGACTCTACACTCAAGACTCACTCATAGG 900
QY 1274 cacaacccctcggaaacaacacacatcgcctccagcctcggcaccctcaagaatgattatagta 1333
| | | | |
Db 901 CACACCCCTCGGGAAACACACATCCGCTCCAGCTCGACCTCGACACTGATATTTAGTA 960
QY 1334 cgaactcgtgaagcagtggtctataacacccaaaaaata tggcaagagccttaataccagctta 1393
| | | | |
Db 961 CGACTGTATGAGCAGTGTATTAACACCAAAAMATGAGCAAGGGCTTATATATCCAGCTTA 1020
QY 1394 agaaagtgtaacagaaggtttgggaattcagacatcactccagaga tgaacatagaatggtc 1453
| | | | |
Db 1021 AGAAAGGTACAGAGAGTTTGGATTCAGCATCTCCAGATGTATACAAATAGGTGGCT 1080
QY 1454 cagctccaatcctatgtgaaaaaacatctcccccggggggcgagccatc tcaaga tggcgcgac 1513
| | | | |
Db 1081 CAGCTCCATCATGTGAAAAAATCTCTCCCGGGGGGGCGGCATTCAGATGGCGGAC 1140
QY 1514 ttaagggcagagacagacttatagaagtaaat tgaagttagattagttgggcaaatcccaag 1573
| | | | |
Db 1141 TTAAAGCGAGGAGACAGACTTTATAGAGTAAATGAGATGATTAGTGGGCAATCCCAAG 1200
QY 1574 aggaagtggtttcgtctgtttgagaagcaccaaga tgaagaagaaacttgagcccttcgtct 1633
| | | | |
Db 1201 AGGAAGTGTTCCTGCTGTGAGAACACCAAGATGGAAGAACTGTGACCTTCCTGCTGT 1260
QY 1634 ttcggcaggaagaacgcttcccaacccaaggaacttgaat tgcagagcagaacccagatgacaga 1693
| | | | |
Db 1261 TTCGCCAGGAAGACGCTTCCACCCAAAGGAACTGAATGACAGCAAGCCAGATGACAGA 1320
QY 1694 ttccaaaagaacagaaagcagaaga tggagatatgtcttcctcacctcgtatgacacagag 1753
| | | | |
Db 1321 TTCCAAAAAGAAACGAAACGAAAGATGAGGATTTCTTTCACCTGATGGCACCAAGG 1380
QY 1754 aattctcagatctgaagtcaccaactaaatgattcagagatc tcaagccttggtgtcagtg 1813
| | | | |
Db 1381 AATTCTGTGATTTGAAGTCCCACTTAGTGAATTCAGATCTCAGAGCCTTGGGTGCAGTG 1440
QY 1814 tcaaaagttaaccggttcaaaagaagaacacgcagatt tgggaa tcttgtcaagtcocatia 1873
| | | | |
Db 1441 TCAAAGGTAAACGGGTCAAAAGAAACCAAGCAGATTTGGGAATCTTTGTCAAGTCCATTA 1500
QY 1874 ttaaatggagagcagcatctaaagatggaaggttcggttgaa tgaatcaactcagatagcag 1933
| | | | |

Db 1501 TTAAATGGAGGACACCATCTAAAGATGAAAGCTTCGGGTGATGATCAACTGATAGCAG 1560
QY 1934 taaatggagaatccctgtttgggcaagacaaaccaaagatccat tggaaacctaaagaagt 1993
| | | | |
Db 1561 TAAATGGAATTCCTGTTGGCAAGACAAACCAAGATGCCATGTGAACCTCTAAGAAAGT 1620
QY 1994 ctaigtctactgaaggaataaagagaatgatccagctta tigtctgaagagaataa 2053
| | | | |
Db 1621 CTATGCTACTGAAGCAATTAACGAGAAATGATCCAGCTTATTTGTGCAAGAGATTA 1680
QY 2054 gcaagtgcgaatgagcttgaagctaaccttggagcccccctggagccctgaagctgcccattgaa 2113
| | | | |
Db 1681 GCAAGTGCATAGCTGAAGTCAACCTGGGAGGCCCTCCGACCTGACCTGAGCTGCATTTGAA 1740
QY 2114 cagcgttgatgataagaagaagaagaatttccattcccttcacagttgagatggagggc 2173
| | | | |
Db 1741 CAGCGTTGGATATATAGAACACGAAATTTCCATTTCCCTTACAGTGGGATTTGAGGGC 1800
QY 2174 ttgatgaatcgcagcaggaataatgctgcctcagtagga taaatgggtgagtcagtaaat 2233
| | | | |
Db 1801 TTGATGAATCGCCACGACAAATGCTGCCCTCAGTAGATATG-----GGTAAAT 1851
QY 2234 accagctgcctcctacagtgaaatagccccaaagatgaac tgcattatagaagatgaca 2293
| | | | |
Db 1852 ACCACCTGCCCTACAGTAATATGCCCAAGATCACACTGTCATTAATAGAAATGAGACA 1911
QY 2294 ggttgcaagtgtcttcctcacatctctcagcagtcctccttcagagctccatgatatg 2353
| | | | |
Db 1912 GGTTCGACAGTCTTCTCCACATCTCTTGACACAGTCTCTTCCTTCACACTCCCATGATAG 1971
QY 2354 tggggttttgtagcaggaatgctgtagtacttggccaaggtctgaatcagtgat ttcagcgg 2413
| | | | |
Db 1972 TGGGTTTGTGACGGCAGATGCTGTACTTGGGCCAAAGCTGCATCACTGATTTGACGGC 2031
QY 2414 actgctctttgtagtcagatgtgtatccagtgcttctgtcttcaagagagaagatttggac 2473
| | | | |
Db 2032 ACTGCTCTTTGAGTCCAGATGTTGATCCAGTTCCTGCTTTCAACCAAGGATTTTGGAC 2091
QY 2474 gtcagatgtagtcagaanaacgacaaagcaatttccagatgcag tcaattgattcgg 2533
| | | | |
Db 2092 GTCAGA----- 2097
QY 2534 ttaaaacacgaataatcaaaaagatgatttagta tgaactgaacgaactaaactcaata 2593
| | | | |
Db 2098 -----CTGACGAGACTTAAACTCAATA 2118
QY 2594 cagtggatgaacgaagaagcaggttctccacgaagagatgt tgggtccttccctgggtctga 2653
| | | | |
Db 2119 CAGTGAATGACCAAGAAACAGAGTTCACACAGAGATGTGGGTCTTCCTGGGTCTGA 2178
QY 2654 agaaagtcagatcgtttggagatctcagacgcaggttc cgaagtgatcttgaatgggg 2713
| | | | |
Db 2179 AAGAGTCAGACTCATTTGGAGAGTCTGCACACCGCACTTCGCGAGGACTTTGAATGGGG 2238
QY 2714 atattccttccatcgttccacgcgcgcgagtaataacagagagcaggtg tgaatgagaagt 2773
| | | | |
Db 2239 ATATTCCTTTCATGCTCCACGGCCGCGATATATCAGAGGACGAGGATCAATGAGAGCT 2298
QY 2774 tcaaggttcctcgcgaataatttatagaataacccgcgttagatgtgtatgtaagagca 2833
| | | | |
Db 2299 TCAGAGCTCCATCGACAAATCTTATGATTAACCCCGGTAGATGATGATGAGAGGCA 2358
QY 2834 tggagaccttggaaagaacacagaaagaat tcaagaatcagaaggaagaagctgtatcca 2893
| | | | |
Db 2359 TGGAGAGCTTGGAGAGAGACAGAGAAAGATTCAAGATCAGAGAGAGAGTCTGTATCCA 2418
QY 2894 cagcagtgatcaagccttcccaactctcgtgagagacaaat tgaatggaacccaagagaag 2953
| | | | |
Db 2419 CAGCCAGTATGATAGCTTCCCACTCTCTGAGAGACAAATGATGATGAACCAAGAGAAAG 2478
QY 2954 gtgataagctatagaanaaagatgaaaactggaaaagaagaagaagatagagata 3013
| | | | |

Db 2479 GTGATAAGACTGATAGAAAAAAGATTAACCTGCAAAAAAGAAAGAAAGATAGAGATA 2538

QY 3014 aggaagaagctaaatgaagaagccaagaaggaatgctgaagggcttggagacatcttca 3073
|||||
Db 2539 AGGAAGAAGATTAATAAAGCAAGAGGAATGCTGAAGGGCTTGGAGACATGTTCA 2598

QY 3074 ggcttgcaaac 3085
| ||| |||||
Db 2599 GCCTTGCCAAAC 2610

RESULT 15
AK027735 2718 bp mRNA linear PRI 15-MAY-2001
LOCUS Homo sapiens CDNA FLJ14829 fis, clone OVARC1000945, moderately
DEFINITION similar to Rattus norvegicus mRNA for atypical PKC specific binding
protein.
ACCESSION AK027735 GI:14042633
VERSION AK027735.1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens ovary, tumor tissue CDNA to mRNA, clone_1lb:OVARC1
clone:OVARC1000945.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Mishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosotani,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,A.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,K.,
Nakanura,Y., Nagahari,K., Masuhio,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human CDNA sequencing project
Unpublished
2 (bases 1 to 2718)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomisehi@co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human CDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; CDNA full insert sequencing;
Research Association for Biotechnology; CDNA library construction,
5'- 6' 3'-end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Source Location/Qualifiers
1..2718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OVARC1000945"
/tissue_type="ovary, tumor tissue"
/clone_1lb="OVARC1"
/note="Cloning vector: pME18SFL3"
33..2630
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55330.1"
/db_xref="GI:14042634"
/translation="MPLAVRRSDPALIGISTVSDNSFSSSEPSRKNPRTWSTAF
LKONTASPRACDRKDENYRSLPRDTSSNNSNQQRNANSSLSASHPMVKWEKOE
QDEGTEDNSRVEPVHADTGLEHINPFLSDMDVAKLVEPNDBGPIGIHVPPSAG
GRTGLVLRLEKGAHEHNLFRFENDCIYRINDGDKLNRFEQAQHMFQAMPTPII
MFHYVPANKEOYEOLDSERKNYSSRFSFSDSOYIDNRSVNSAGLHYORAPLNP
PEQDSHSRLPHASHSGKSPASAPASAPVVESTTVSSGNTKIGRLNIOKKGE
GLGFSITSRVYTTGGSPITYKNTLPKCAIIOBRRLAKGDRLLIVNVDLVKQOEY
VSLRSTRKMGTVSLVLFROEDAFHRELAEPSOMIIPETKAEDEDVILTPDGTIE
FLTFEVLSDSGSLGVSKNSKKNHNDLGIFFVSIINGGAASDGRILRVNDOLI
AVNCSLIGTNDAMETLRSMSSTENKGNKDLIYARISKTSEKNTKSPSPGPEL
PIETALDRERIRISHLSYSGIEGDESPRNMAOLIRLMGKTKSLPEYVMODDVILE
DDRILPVLPPLHSDSSSHDDVGFYADACTAKAKAISDSADCSLSLDPDVPVLAQFR

EGFGQDDETKLINTVDDOKAGSPSDVGPISLCKKSSSIESIQTVAEYTLNGDIPF
RPRPRIIRGRCNESFPRAIDKSYDKPAVDDDEMEETLEEDTESSSRGRESTAS
DQPSHLEROMNGNOEKGDTRKDKTKERKKRDKREKMKAKKMLKGLGDMFS
LAKLPERK"

BASE COUNT 840 a 586 c 688 g 604 t
ORIGIN

Query Match 43.5%; Score 2476.8; DB 9; Length 2718;
Best Local Similarity 96.0%; Pred. No. 0;
Matches 2603; Conservative 0; Mismatches 7; Indels 102; Gaps 2;

QY 374 ttgaagtcacacctaagtccttcctgcagcaaatctgctctcatgtctgcagcagtagt 433
Db 1 TTGAGGTACACACCTTCAGTCTTCCTTGAGCAAAATATGCTTCTCATGTTCAGCAGCATAGTG 60

QY 434 accagctcaattggtgctctcctcactctctgtaagtagtaatttttctctgaagc 493
Db 61 ACCAGCTCAATTGGCGCTTCCACCTTCTCAGTAGTATGTAATTTTCTCTGAAAGC 120

QY 494 ctcaaggaanaaaatccacacacgctgltcaacaacagctggtcttctcctcaagcagaacatg 553
Db 121 CTCACAGGAAAAATCCACACAGCTGCTCAGCAACAACAGCTGGCTTCTTCAAGCAGAACACTG 180

QY 554 ctggagctcctaacaacctgcgacaagaagaagatgaataaactacagaagctcccgagg 613
Db 181 CTGGAGTCTCTTAAAGCTCTGACAGGAAGAAATGAATAACTACAGAAAGCTCCCGGG 240

QY 614 atactagtaactgcttcaacaaattcagagagacaatgctcgtctgctcagtgcca 673
Db 241 ATACTAGTAAGTGGTCTTACCAATTTTACAGAGACATCTCTCGTCTCTAGTGTCCA 300

QY 674 gtccaccaatgltggygaagtgtcgtggaagaagaacaagaagatgagatggagaag 733
Db 301 GTCACCCAAATGTTGGGCAAGTGCAGAGAAACAAAGAAAGATGAGATGGGACAGAG 360

QY 734 aggaatacagtcgtgttggaacctgtgacatgctgcacaggggtttagcatatacca 793
Db 361 AGGATTAACAGTGGTGAACCTGTTGACATGCTGACAGGGGTTTGAGCATATACCA 420

QY 794 accttctcgtatgatalgtlaaagctcgtlaagaagtcctccaaagatgagggccttg 853
Db 421 ACTTTCTGTGATGATATGTAAGCTCGTAAGATCCCAACAGATGAGGGCTCTGG 480

QY 854 gaatcatalgtagtgccttcaagtgctcgtgagcgagcaaaccttgggttattagtaaac 913
Db 481 GAATCATTGATGCTCTTTCAGTCTGAGCGGCGCAGAACCTTGGGGTATTAGTAAC 540

QY 914 gattggagaagtggtgaagctgaacatgaataatcttttcgtggaatgattgcatg 973
Db 541 GATTGGAGAAAGGTGTAAAGCTGAACATGAAATCTTTTCTGTGAATATGTCATTG 600

QY 974 tcaagatlaatgacgacgaccttcgaataagaagatttgaacaagacacaatalgttcc 1033
Db 601 TCAGGATTAATGATGCGGACCTTGAAATAGAAATTTGAACAAGCACAACATATGTTTC 660

QY 1034 gccaaagccatgctgaacccaatcattggttccatgtgttccctgcgacgaataaagac 1093
Db 661 GCCAAGGCATCGTGAACCCATCTTGTGTCATGTGGTTCGTGCACAATTAAGAGC 720

QY 1094 agtatgaacaactatcccaagttagaagaacaatttactatccaagccgttttagccctg 1153
Db 721 AGTATGAACAACCTTCCCAAGTGAAGAAGAACATTAATATTCAAGCGGTTTTCAGCCTG 780

QY 1154 acaagccaglatatgacacacagagtgtaaacagltgcagggtctcaacggttgcagagag 1213
Db 781 ACAGCAAGTATATTGACAAACAGAGTGTGAACACTGACAGGCTTACACAGGTGACAGAG 840

QY 1214 caaccggactgaacaccccgctgagcagatagaactctcactcaagacttactactagcg 1273
Db 841 CACCCGCACTAACCCACCGGCTGAGATGAGACTCTCACTCAAGACTACTCAATACG 900

QY 1274 caaacccctcggaanaacacacatcgcgtccagcctcgacacatcagaatgtattagta 1333

```
Db 901 CACACCCCTCGGGAACACCATCCGCTCCAGCTCGGACCTCAGATGATTTAGTA 960
Oy 1334 cgaactgaagcagtggtatataacacacaaataagcaaggcttaataccagctta 1393
Db 961 CGACTGTAGAGCAGTGGTTATATACACCAAAAAATAGCAGAGGCTTATATCCAGCTTA 1020
Oy 1334 agaaagtgacagaagaggtttggatcagcaatcaactccaagagatgaataagtgct 1453
Db 1021 AGAAAAGTACAGAAAGTTGGATTGACATCCTCCAGAGATGATTAACAATAGCTGGCT 1080
Oy 1454 cagctccaatctatgtgaaaaaatctctcccccggggggccgagcttaagatggcgac 1513
Db 1081 CAGCTCCATCTATGTGAAAAACATTTCTCCCGGGGGGCGGCAATCAGATGGCGGAC 1140
Oy 1514 ttaagcgagagacagactatagagtgtaagtgaatgaatgaatgaatgaatgaatgaat 1573
Db 1141 TTAAGGCAAGACAGACTTATAGAGGTAATAGAGTAGATTTATGGGCAAAATCCCAAG 1200
Oy 1574 aggaagttgttgcgtgttgagaagacccaagaatgaagaagatgtgagccttgcgtct 1633
Db 1201 AGGAAGTTGTTTCCGCTGTGAGAACCCAGATGGAAGAACTGTGAGCCTTCTGCTCT 1260
Oy 1634 ttccgcaagaaagagccttcccccgaaggaaatgaatgaagaagcaagcagatgacaa 1693
Db 1261 TTCCGCAAGAAACGCTTCCACCCAGAGGAACTGAAATGCAAGCCAGATGCAAGA 1320
Oy 1694 ttccaaagaaacgaagacgaagaatgaagatattgttcttaacatgaatggcacaag 1753
Db 1321 TTCAAAAGAAAGAAAGAGAGATGAGATATTGTTCTTACACTGATGGCACACAGG 1380
Oy 1754 aattctcaatctgaagtcaccaatgaatgaatgaatgaatgaatgaatgaatgaatga 1813
Db 1381 AATTCTCAATCAATTTAAATCCCACTTAGATGAGATGCAAGCTCAGGCTTGTGTCAGTG 1440
Oy 1814 tcaaaagtaaacggtcacaagaagaacacagcagatttggaatcttctgcaagtcacaa 1873
Db 1441 TCAAAAGTAACCGGTCCAAAAGAAACACGCGATTTGGGAATCTTTGTCAAAGTCCATTA 1500
Oy 1874 ttaatgagagacagcatctaaagaatgaagagcttcgggtgaatgaatgaatgaatga 1933
Db 1501 TTAATGAGAGACAGCATCTAAAGATGAGAGCTTCGGGTGAATGATCAACGATGACAG 1560
Oy 1934 taatgagagaaatccctgttggcagaagaacaaagaatgcaatgaagaagct 1993
Db 1561 TAAAGTGAGAAATCCCTGTTGGGCAAGACAAACCAAGATGCCATGGAACCCCTAAGAAAGGT 1620
Oy 1994 ctaigtctactgaaggaataaaggaatgaatgaatgaatgaatgaatgaatgaatga 2053
Db 1621 CTATGCTTACTGAAAGCAATTAACGAGGATATCCAGCTTATGTTGCAAGGAGATTA 1680
Oy 2054 gcaagtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 2113
Db 1681 GCAAGTGAATGAGCTGAAGTCACTCGGAGGCCCTCTGAGCTGAGCTGACCATTTGAAA 1740
Oy 2114 cagcgttggatgaatgaagaagaatgaatgaatgaatgaatgaatgaatgaatgaatga 2173
Db 1741 CAGCCTTGGATGATGAGAGAAACGAAGAAATTCATCCCTTACAGATGAGATGAGAGGCG 1800
Oy 2174 ttgatgaatcgcccaagcaaaatgtctgcacctgaatgaatgaatgaatgaatgaatga 2233
Db 1801 TTGATGAATCGCCCAAGCAAAATGCTGCTTCATGATGATTAATG-----GGTAAAT 1851
Oy 2234 accagctgtccctcaagtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 2293
Db 1852 ACCAGCTGTCCCTACAGAGATATGCCCCAAGATGACACACTGTCATTTATGAAGATGACA 1911
Oy 2294 ggttgcagatgcttctcccaatctctgacagatcctctccagctcccatgatgatg 2353
Db 1912 GGTTCGCAATGCTTCCCTCACATCTCTTGACACAGCTCTTCCAGCTCCCATGATGATG 1971
Oy 2354 tggggttttgacgagatgtgtaacttggccaaggcttgaatcagttgaatgaatgaatga 2413
|||||
```

```
Db 1972 TGGGTTTGTAGCGGACATGCTGTGACTTGGGCCCAAGCTCAGATGATTCAGCGG 2031
Oy 2414 actgctcttgagtcagagatgttgatccagttcttgccttcaacgaagaagatttgac 2473
Db 2032 ACTGCTCTTGTAGTCCAGATGTTGATCCAGATGTTGCTTGTTCATCAAGAGAGATTTGGAC 2091
Oy 2474 gtcaagatgtgcagaaaaacgacaaagaatlttcaagatgcagttcaatgtgattcg 2533
Db 2092 GTGAGA----- 2097
Oy 2534 ttaaacacgaataatcaaaaagcatgattagtagtagtagtagtagtagtagtagtagtag 2593
Db 2098 -----CTACGAGATTAATCAATA 2118
Oy 2594 cagtgtatgaccagaagaagcaggttctccacagagatgttggtctctccctgggtctga 2653
Db 2119 CAGTGGATGACGAGAAAGCAGGTTTCCAGAGAGATGTTGGTCTTCCCGGGGTCTGA 2178
Oy 2654 agaaagtaagctcgtttgagagatctgcagaccgcaagttgcgaagtgacttgaatggg 2713
Db 2179 AGAAGTCAAGCTCATTTGGAGAGTCTGCAGACCGCAGTTGCCAGAGTGACTTGAATGGGG 2238
Oy 2714 atattccttccatcgttcacgagccgagtaatacgaagaagcagagatgaatgaatgaatga 2773
Db 2239 ATATTCTTTCATCTCTCCACGCGCGGATTAATCAGAGCGAGGAGATCAATGAGAGCT 2298
Oy 2774 tcaagctgcacatgcacaatcttataaaccggtgaatgaatgaatgaatgaatgaatga 2833
Db 2299 TCAGAGCTCCATGCAATATCTTATGATTAACCCCGGTAGATGATGATGATGAAGCA 2358
Oy 2834 tggagaccttggagaagaacacagaagaagaatgaatgaatgaatgaatgaatgaatga 2893
Db 2359 TGGAGACCTTGGAGAAAGACACAGAAAGAAATTCAGAGATCAGGAGAGAGTGTATCCA 2418
Oy 2894 cagcagatgacagccttccacactctctgagagacaaatgaatgaatgaatgaatgaatga 2953
Db 2419 CAGCCAGTATATAGCCTTCCACTCTCTGAGAGACAAATGAAATGAAACCAAGAGAAAG 2478
Oy 2954 gtgataagaatgataaagaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 3013
Db 2479 GTGATTAAGACTGATGAAAAAAGATTAATGAAAAAGAAAAAGAAAGATGAGATA 2538
Oy 3014 aggaagaagataaatagaagaagcaagaagaagaatgtgaaggcttggagagacatgttca 3073
Db 2539 AGGAGAAAGATTAATAAGAAAGCAAGCAAGGATGCTGAAGGCTTGGAGAGACATGTTCA 2598
Oy 3074 ggttggcaaac 3085
|||||
Db 2599 GCCTTGCCAAAC 2610
```

Search completed: July 24, 2002, 11:12:39
Job time: 16331 sec

Thu Jul 25 08:38:18 2002

OM of: US-09-757-781-2 to: GenEmbl:* out_format : pfs

Date: Jul 24, 2002 6:10 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+2.p2n.model -DEV=x1h
 -O=/cgn2.1/USPTO.spool/US09757781/Runat_22072002.154317_18590/app_query.fasta.1.1434
 -DB=GenEmbl -OPMT=fastap -SUFFIX=olig.rge -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPECL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -DELTOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAEXT=60.000
 -DELTOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NOM=ext
 -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US09757781.ecgn1.1.5806 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-757-781-2
 Query length: 1356
 Database: GenEmbl:*
 Database sequences: 1797656
 Database length: 187333701
 Search time (sec): 3888.640000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
 WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	EScore	Len	Documentation
gb.pr.AF67002	+ 1356.00	25804.76	0.0	4071 AF67002 Homo sapiens partit
gb.pr.AF67003	+ 1113.00	21171.89	0.0	4062 AF67003 Homo sapiens partit
gb.pr.AF196185	+ 1113.00	21169.30	0.0	5958 AF196185 Homo sapiens atypica
gb.pr.AF67004	+ 1025.00	19494.31	0.0	3960 AF67004 Homo sapiens partit
gb.pr.AF196186	+ 781.00	14843.26	0.0	3470 AF196186 Homo sapiens atypica
gb.pr.AK024668	+ 614.00	11662.69	0.0	2116 AK024668 Homo sapiens cDNA: H
gb.pr.AF332592	+ 556.00	10554.61	0.0	2867 AF332592 Homo sapiens atypica
gb.pr.AF332593	+ 556.00	10552.90	0.0	3822 AF332593 Homo sapiens atypica
gb.pr.AF67006	+ 556.00	10552.70	0.0	3933 AF67006 Homo sapiens partit
gb.pr.AF525293	+ 500.00	9485.27	0.0	3801 AF525293 Homo sapiens partit
gb.pr.BC011711	+ 497.00	9433.97	0.0	1592 BC011711 Homo sapiens clone M
gb.pr.AF172288	+ 427.00	8097.84	0.0	2004 AF172288 Homo sapiens cDNA: H
gb.pr.BD004905	+ 420.00	7962.32	0.0	2718 BD004905 Homo sapiens cDNA: H
gb.pr.AK027735	+ 356.00	6742.19	0.0	2696 AK027735 Homo sapiens cDNA: H
gb.pr.AK000761	+ 318.00	6015.50	0.0	3735 AK000761 Homo sapiens cDNA: H
gb.pr.AK025892	+ 255.00	4819.43	4.4e-260	1772 AK025892 Homo sapiens cDNA: H
gb.pr.AK025892	+ 131.00	2464.99	6.1e-129	425 AK025892 Homo sapiens cDNA: H
gb.pr.AK025892	+ 131.00	2427.61	7.4e-127	106198 AK025892 Homo sapiens cDNA: H
gb.pr.AK025892	+ 131.00	2425.80	9.4e-127	138688 AK025892 Homo sapiens cDNA: H
gb.pr.AK025892	+ 131.00	2423.44	1.3e-126	196374 AK025892 Homo sapiens cDNA: H
gb.pr.AK025892	+ 129.00	2389.59	9.7e-125	104357 AK025892 Homo sapiens cDNA: H
gb.pr.AK025892	+ 102.00	1885.29	1.2e-96	22279 U82210 Homo sapiens chromos
gb.pr.AK025892	+ 89.00	1661.95	3.3e-84	597 AK025892 Homo sapiens chromos
gb.pr.AK025892	+ 88.00	1627.84	2.6e-82	5500 AK025892 Homo sapiens chromos
gb.pr.AK025892	+ 84.00	1569.33	4.7e-79	400 AK025892 Homo sapiens chromos
gb.pr.AK025892	+ 81.00	1471.27	1.4e-73	165146 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 81.00	1471.25	1.4e-73	167163 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 81.00	1471.18	1.4e-73	169337 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 81.00	1470.29	1.4e-73	193131 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 78.00	1452.91	1.5e-72	540 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 77.00	1420.91	9.9e-71	4167 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 60.00	1073.74	1.9e-51	109816 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 60.00	1072.18	2.3e-51	138322 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 60.00	1069.81	3.2e-51	196259 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 52.00	918.33	8.7e-43	168213 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 44.00	808.55	1.1e-36	305 AK025892 Homo sapiens clone
gb.pr.AK025892	+ 44.00	808.47	1.1e-36	309 AK025892 Homo sapiens clone

seq_name: gb.pr.AF67002

seq_documentation_block:

seq_name	gb.pr.AF67002	4071 bp	MRNA	linear	PRI	7-FEB-2002
gb.pr.AF67002	-	40.00	689.31	4.9e-30	173987	1 AL450337 Human DNA sequen
gb.pr.AF67002	-	38.00	651.45	6.3e-28	167163	1 AC013705 Homo sapiens cto
gb.pr.AF67002	-	34.00	614.50	7.3e-26	504	1 AL185634 Sequence 1329 from
gb.pr.AF67002	-	34.00	574.83	1.2e-23	176405	1 AL185634 Human DNA sequen
gb.pr.AF67002	-	33.00	562.01	6.1e-23	70087	1 AC105981 Mus musculus clon
gb.pr.AF67002	-	28.00	466.69	1.2e-17	70087	1 AC105981 Mus musculus clon

accession

AF67002.1 GI:18568343

version

human.

source

human.

organism

human.

reference

human.

authors

human.

journal

human.

features

human.

source

human.

gene

human.

cds

human.

misc-feature

human.

misc-feature

human.

misc-feature

human.

misc-feature

human.

misc-feature

human.

misc-feature

human.

```

misc_feature      /note="Region: PDZ domain"
                  /gene=".1638"
misc_feature      /note="Region: PDZ domain"
                  /gene=".2028"
BASE COUNT       1220 a 923 c 1079 g 849 t
ORIGIN

Alignment_scores:
  Quality: 1356.00      Length: 1356
  Ratio: 1.000          Gaps: 0
  Alignment_block:      Percent Identity: 100.000
US-09-757-781-2 x AF467002 ..

Align seg 1/1 to: AF467002 from: 1 to: 4071

1 MetLysValThrValCysPheGlyArgThrArgValValProCysgl 17
  |||||
1 ATGAAGTGAACCGTGTCTCGACGAGACCGGGTGGTGGTCCGCGCGG 50
  |||||
17 YASGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
  |||||
51 GGAGGCGACATGAAGATTTCAGCTCATTCACGACGCGGCGTGCACCCGCT 100
  |||||
34 YHArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnAlaHisArg 50
  |||||
101 ACCGGAAAGCCATCGCCAGATGCCAAACTACTGGATACAGTGCATCGC 150
  |||||
51 LeuGlnHisGlyAspGlyIleLeuAspLeuAspAspIleLeuCysAs 67
  |||||
151 TTGGAAATGAGATGAGAGAAATACACCTTCATGACATCTTTGTGCA 200
  |||||
67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGluAlaAsp 84
  |||||
201 TGTAGCAGACGATAAAGACAGACTGTRACAGTGTGTGATGAGACAGANTC 250
  |||||
84 roHisHisGlyLysPGLYThrSerAlaSerSerThrGlyThrGlnSer 100
  |||||
251 CACATACAGGAGGTGATGGCCACGATGCCAGTTCACGCGATACCCAGAGC 300
  |||||
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117
  |||||
301 CCAGAGATATTTGGTGTAGTGGCTTGCGACCAACATGCTCAGCTTTTCA 350
  |||||
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
  |||||
351 GCCTTACCAAGCAACAGTGAATTTAGGTCAACCTTCAGTCTTCGAG 400
  |||||
134 IAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
  |||||
401 CAATATGCTCTTCATGTTCGACGAGTACACCCAGCTCTAATTGGC 450
  |||||
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerArg 167
  |||||
451 CTCTCCACTCTGTCAGTGAATGTAATTTTCCTGTAAGAGCCCTCAAG 500
  |||||
167 gLysAsnProThrArgTTPSerThrThrAlaGlyPheLeuLysGlnAsnT 184
  |||||
501 GAAAAATCCACACGCTGCTCAACACAGCTGGCTTCTTACAGAACACA 550
  |||||
184 hAlaIleSerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
  |||||
551 CTGCTGGAGAGTCTTAAACCTGGACACGGAAGAAAGTGAACAACTACACA 600
  |||||
201 SerLeuProArgAspThrSerAsnTTPSerAsnGlnPheGlnAlaGAspAs 217
  |||||
601 AGCCTCCGCGGAGATACGTAAGTAACTGCTTACCACTTTCAGAGAGACAA 650

217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
  |||||
651 TGTCTGCTGCTCTCTGAGTGCAGAGTACCCCAATGGTGGCAAGTGGCTGG 700
  |||||
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
  |||||
701 AGAAACAAAGAACAGATGAGATGGGACAGACAGAGAGATTAACAGCGTGT 750
  |||||
251 GluProValGlyHisAlaAspThrClyLeuGluHisIleProAsnPheSe 267
  |||||
751 GAACCTGTTGACATGCTGACACGAGGTTTGGAGCATATACCAACTTTTC 800
  |||||
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
  |||||
801 TCTGATGATATGTAAGCTCTAGAAAGTCCCAACGATGGAGGCGCTC 850
  |||||
284 euGlyIleHisValAlaProPheSerAlaArgGlyLysArgThrLeuGly 300
  |||||
851 TGGGAATCCATGATGTCCTTTCTAGTCTCGAGGCGGACGAACCTGGGG 900
  |||||
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisIleGlnAsnLe 317
  |||||
901 TTATTAGTAAACGATTTGAGAAAGGTGTAAAGCTGAACATGAAATCT 950
  |||||
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
  |||||
951 TTTTCTGTAGATGATTTCTCATTTGACAGATTATATGATGCGACCTTCA 1000
  |||||
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
  |||||
1001 ATAGAAGATTTGAACAAAGCACAAATATGTTCCGCAAGCCATCCGTACA 1050
  |||||
351 ProIleIleTrpPheHisValAlaProAlaAlaAsnLysGluGlnTyrG 367
  |||||
1051 CCCATCTATTGGTTCATGCTGCTCTCCACGAATTAAGACAGATATA 1100
  |||||
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
  |||||
1101 ACNACTATCCCAAGTGAAGAAACAAATTAATCTATTCACGCTTTTACCC 1150
  |||||
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
  |||||
1151 CTGACAGCCACTATATGACACACAGAGGTGTACACAGTGCAGGCTTAC 1200
  |||||
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
  |||||
1201 ACGGTGAGAGAGACCCCGACTGAACACCCCGCTGAGCAGATACACTC 1250
  |||||
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
  |||||
1251 TCACCTAAGACATACCTCATACGCGACACCCCTGGGAAAAACCAATCCG 1300
  |||||
434 lProAlaSerAlaProGluAsnValPheSerThrThrValSerSerGly 450
  |||||
1301 CTCCAGCCCTCGCACCTCAAGATGTAATTAGTACACTGTAAAGCAGTGGT 1350
  |||||
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
  |||||
1351 TATAACCAAAAAAATATAGCAAGAGCTTAATATCCAGGTAAAGAAAG 1400
  |||||
467 YThrGlnGlyLysGlyPheSerIleThrSerArgAspValThrIleGlyG 484
  |||||
1401 TACAGAAAGTTGGGATTCACATCACCCTTCCAAAGATGTAAACAAATAGT 1450
  |||||
484 lYserAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
  |||||
1451 GCTCAGCTCCCAATCATGTGAAAAACATTTCTCCCGGGGCGGCCCATTT 1500
  |||||
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
  |||||
1501 CAGGATGCGGACTTAAAGCAGAGACAGACTTATAAAGTAAATGAGAGT 1550
  |||||
517 lAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534

```

```
|||||
1551 AGATTAGTGGGCAATCCCAAGAGAAAGTTGTCGCTGTGAGAACGA 1600
534 hrlsMetGluGlyThrValSerLeuLeuValPheArgGlnIuAspAla 550
|||||
1601 CCAAGATGGAAGAACTGTGAGCCTTGTGCTCTTCCCGAGAAAGAGGCC 1650
551 PheHisProArgIuLeuAsnAlaGluProSerGlnMetGlnIleProly 567
|||||
1651 TTCACCCCAAGGAACTGAAATGACAGCCAAAGCCAGATGCAATGCCAA 1700
567 scGluThrLysAlaGluAspGluAspIleValIleuThrProAspGlyThra 584
|||||
1701 AGAAGCAAGAAAGCAGAAAGATCAGATATGTTCTTACACCTGATGGCACCA 1750
584 rglIuPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
|||||
1751 GGAATTTTCGACATTTGAAATGCCACTTAAATATTCAGATCTCGAGGC 1800
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspIle 617
|||||
1801 CTGCGTGTCAAGTGTCAAGGTAACCGGTCAAAAGAACCAACGACAGATT 1850
617 uGlyIlePheValLysSerIleIleIleAsnGlyGlyAlaAlaSerLysAsp 634
|||||
1851 GGAATCTTTGTCAGTCCATTTATTAATGAGAGAGCAGCATCTAAAGATG 1900
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyIuSerLeu 650
|||||
1901 GAAGGCTTCGGGTGAATGATCAACTGATAGCACTAAATGAGAAATCCCTG 1950
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
|||||
1951 TTGGGCAAGCAAAACCAAGATGCCATGGAAACCTTAAGAGGCTATATGTC 2000
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
|||||
2001 TACTGAAGCAATAAACGAGAAATGATCCAGCTATTGTTGCAAGAGAA 2050
684 IeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700
|||||
2051 TAAGCAAGTCGAATGAGCTGCAACTCACTGGAGCCCCCTGGACCTGAG 2100
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
|||||
2101 CTGCGCATTTGAAACAGCGTTGATGATAGAGAACGAAGATTTCCCATTC 2150
717 rLeuTyrrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
|||||
2151 CCTCTACAGTGGGATGAGGGGCTGATGAAATCCCCCAGCAAGAAATGCTG 2200
734 IeLeuSerArgIleMetGlyGluSerGlyLysTyrrGlnLeuSerProThr 750
|||||
2201 CCTCTACAGTGAATGAGGAGTCAAGCTCAATACCAAGCTGCTCCCTACA 2250
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
|||||
2251 GTGAATATATGCCCAAGATGACACTGTCATTATGAAAGATGACAGGTTGCC 2300
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerSerHisAsp 784
|||||
2301 AGTGTCTCTCCACATCTCTGTACCAAGTCCCTTCACAGCTCCCATGATG 2350
784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800
|||||
2351 ATGTGGGGTTTGTGACGCGAGATGCTGTACTTGGGCCAAAGGCTGCATATC 2400
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
2401 AGTGATTACAGCCGACTGCTCTTGTAGTCCAGATGTTGATCCAGTCTTGTGC 2450
817 aPheGlnArgGluGlyPheGlyLysArgIuSerMetSerGlyLysArgThrL 834
|||||
```

```
2451 TTTTCAACGAGAAAGATTGGAGCTCAGAGTANGTCAGAAAAACGACAA 2500
834 YsGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
|||||
2501 AGCAATTTTCAGATGCCGATCAATTTGGATTTTCGTTAAACACGAAATACA 2550
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
|||||
2551 AAAAGCATGTGATTTAGTATAGTGAAGAGAGACTAAACTCAATACAGTGA 2600
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
|||||
2601 TGACCAAGAAAGCAGGTTCTCCACAGAGATGTGGCTCTTCCCTGGGTC 2650
884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
|||||
2651 TGAAGAAAGTCAAGCTCGTTGGAGAGTCTGCAGACCGCAAGTTGCCGAGGTG 2700
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917
|||||
2701 ACTTGAATGGGATATTCCTTTCATCTGTCACAGCGCGGAGATTAATCAG 2750
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyra 934
|||||
2751 AGGCAAGGATGCATATGATGAGCTTCAGAGCTGCCATGCACAAATCTTATG 2800
934 sPlysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950
|||||
2801 ATTAACCCCGGGGTGATGATGATGATCAAGGATGGAGACCTTGGAAAGAA 2850
951 AspThrGlnLysSerSerArgSerGlyLysArgLysValSerThrAlaSe 967
|||||
2851 GACACAGAAAGAAAGTTCAAGATCAGGAGAGAGTCTGTATCCACAGCCAG 2900
967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyLysAsnGlnL 984
|||||
2901 TGATCAGCCCTTCCACCTCTCTGGAGAGACAAATGAAGAAACCAAGAGA 2950
984 YsGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
|||||
2951 AAGGTGATTAAGACTGATAGAAAAGATTAAGAACTGGAAAAAGAAAGAG 3000
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
|||||
3001 AAAGATGAGATTAAGAGAAAGATTAAGAAATGAAGCAAGAAAGGAAATGCT 3050
1017 uLysGlyLeuGlyLysMetPheArgPheGlyLysHisArgLysAspAspL 1034
|||||
3051 GAAAGGCTTGGAGACATGTTCAGGTTTGGCAAAACATCGAAAGATGACA 3100
1034 YsIleGluLysThrGlyLysIleLysIleGlnLysSerPheThrSerGlu 1050
|||||
3101 AGATTGGAAGAACCGGTAAATATAAATACAGAAATCTTTACTACATCGAA 3150
1051 GluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysThr 1067
|||||
3151 GAGAGAGAGATACGAATGAAGAGAGAGAGAGAGATTCGAAGCAAAAC 3200
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyraAlaGluIleG 1084
|||||
3201 TCGAGAATTTAGGGAACGACAAAGCTCGAGAGCGGACTATGCTGAAATTC 3250
1084 IAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyrrGlyGly 1100
|||||
3251 AAAGATTTTCATCGGACATTTGGCTGTGATGATAGATTAAATGATAGGGGA 3300
1101 ValSerSerTyrrGluGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
|||||
3301 GTTCTCTTATGAAAGGTTCCATGGGCTCTCAAGCTAGACTCAGAGGCC 3350
1117 oArgGluGlyHisMetLeuAspAlaLeuTyrrAlaGlnValLysLysPro 1134
|||||
3351 ACGAGAAAGGCATATGATGATGCTTGTGTATGCCCAAGTCAAGAAAGCCGC 3400
```

[illegible]

17 yAspGlyHisMetLysValPheSerLeuIleGlnIleAlaValThrArgT 34
 |||||
 51 GGAGCGGCACATGAAGTTTTCAGCCTTCATCCAGACGGGTGAACCCCT 100
 34 yTrArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
 |||||
 101 ACCGGAGGCGCATCGCCAAAGATCCAAACTGATACAGTGGTGCATCGC 150
 51 LeuGlnHisGlyAspGlyLysIleLeuAsnLysAspPheLeuCysAs 67
 |||||
 151 TTGGACATGGAGATGGAGGATACTAGACCTTGATGACATTCCTTGTGA 200
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlyLysAsp 84
 |||||
 201 TGTAGCAGACGATTAAGACACACTGCTAGCAGTGTTCATGACGACGATC 250
 84 roHisHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
 |||||
 251 CACATCACGGAGGTGATGGCACACAGTGCAGTCCACGGGTACCAGAC 300
 101 ProGluIlePheGlySerGlyLeuGlyThrAsnAsnValSerAlaPheG 117
 |||||
 301 CCAGAGATATTGGTAGTGAAGCTTGACCAACATGTTCTCAGCCTTCA 350
 117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
 |||||
 351 GCCTTACCACACCAACAGTGAATTTGAGGTACACACTTCAGTCCGTGAC 400
 134 LAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
 |||||
 401 CAAATATGCCCTTCATGTTGACGACAGTACTGACCCAGCTCTAATTTGC 450
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGlyLysProSerAr 167
 |||||
 451 CTCTCCACTCTGTGACGATGATGATTTTCTCTGAAGAGCCTTCAG 500
 167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
 |||||
 501 GAAAAATCCCAACAGCTGTGCACACACAGCTGCTCTCAAGCAGACAA 550
 184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
 |||||
 551 CTGCTGGAGAGTCTTAAACCTGCGACAGAGAAAGATGAATAACTACAGA 600
 201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
 |||||
 601 AGCCTCCCGGGGATACTAGTACTGCTTAACCAATTTCAAGAGACAA 650
 217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
 |||||
 651 TGTGCGGCGCTCTGAGTGCACACCAATGATGGCGCAAGTGGCGGG 700
 234 LysGlnGlnGluAspGlyLysArgLysGluLysAspAsnSerArgVal 250
 |||||
 701 AGAAACAGAAACAGATGAGATGGACAGAAAGATTAACAGTCTGCTTT 750
 251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
 |||||
 751 GAACCTGTGGACATGCTGACACAGGGTTTGGAGCATATACCCCAACTTTC 800
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
 |||||
 801 TCTGATGATATGTGTAAGCTCGTAGAAGTCCCAACGATGGAGGCGCTC 850
 284 euGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300
 |||||
 851 TGGGAATCCAGTAGTGCCTTTCAGTGTCTGAGGGGCGAGAAACCTGGGG 900
 301 LeuLeuValLysArgLeuGlyLysGlyLysLysAlaGlnHisGlyLysLe 317
 |||||
 901 TTATTAGTAAACGATTGGAGAAAGTGTGTAAGCTGACATGAAATCT 950

317 nPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
 |||||
 951 TTTTGTGAGAAATGATTCATGTGTCAGATTAATGATGGCGACCTTCGAA 1000
 334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgTrp 350
 |||||
 1001 ATAGAAGATTGGAAACAGACAAACATATGTTGGCCAAACCATGTGGTACA 1050
 351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGlnTyrG 367
 |||||
 1051 CCCATCATTTTGGTTCATGTGCTTCTCTCGACGCAATTAAGACAGATACA 1100
 367 uGlnLeuSerGlnSerGlyLysAsnAsnTyrTyrSerSerArgPheSerP 384
 |||||
 1101 ACAACTATCCCAAGTGAAGAACACAAATTAATTAACCCGCTTTTATGCC 1150
 384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
 |||||
 1151 CTGACAGCCAGTATATTGCAACAGAGGTGTGAACAGTGCACAGGCTTCAC 1200
 401 ThrValGlnArgAlaProArgLeuAsnHisProProGlnIleAspSe 417
 |||||
 1201 ACGGTGCAGAGACACCCCGACTGAACCCCGCTGACAGATAGACTTC 1250
 417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
 |||||
 1251 TCACTCAGACATACCTCATAGCCGACACACCCCTCGGGAACCAACCATCCG 1300
 434 LProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
 |||||
 1301 CTCACAGCTCGGACCTCGAATGTATTAGTACGACTGTAAGCATGTGT 1350
 451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysG 467
 |||||
 1351 TATTAACACCAAAAATAGCGCAAGAGCTTAATATCCACTTAAGAAAGC 1400
 467 yThrGlnGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
 |||||
 1401 TACAGAGGTTTGGGATTCAGCATCCTTCAGAGATGTAAACATATGCTG 1450
 484 LysSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
 |||||
 1451 GCTCAGCTCCAAATCTATGAAAAAACATCTCCCCGGGGGGCGGCATT 1500
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGly 517
 |||||
 1501 CAGGATGGCCGCACTTAAGCGACAGACAGACTTAATAGAGTAAATGAGAT 1550
 517 LAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534
 |||||
 1551 AGATTATTAGTGGCAAAATCCCAAGAGGAATGTGTTCCGCTGTGAGAGCA 1600
 534 hTrLysMetGluGlyThrValSerLeuLeuValPheArgGlnLysAla 550
 |||||
 1601 CCAAATGGAAGAACTGTGAGCTTGTGCTTTCGCGCAGAGAGCGGC 1650
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProL 567
 |||||
 1651 TTCCACCCCAAGGAACCTGAATGCACAGCCCAACCCAGATCAATTCCAA 1700
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
 |||||
 1701 AGAAACGAAGACAGAAAGATGAGCATATTGTCTTAACACTGATGGACCA 1750
 584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
 |||||
 1751 GGAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGAGATCTGAGGC 1800
 601 LeuGlyLysSerValLysGlyLysAsnArgSerLysGluAsnHisAlaAsp 617
 |||||
 1801 CTTGGTGTACGTCAAAAGGTAAACCGGTCAAAAGCAACACAGCAGATTT 1850
 617 uGlyIlePheValLysSerIleIleAsnGlyLysAlaAlaSerLysAspG 634

```

1851 GGGAAATCTTGTCAAGTCCATTATATGAGAGACGACATCTTAAGATG 1900
634 1AaGLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLeuSerLeu 650
1901 GAAGGCTTCGGTGAATGATCACTCACTGATAGCAATGAAATGCAATCCCTG 1950
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
1951 TTGGGCAAGACAAACCAAGATGCTCATGGAACCCCTAAGAGGTCTATGTC 2000
667 rThrGluGlyAsnLysArgGlyMetGlnLeuIleValAlaArgArg 684
2001 TACTGAGGCAATAAACGAGATGATCTCAGCTTATTTGTCGAAGAGAA 2050
684 LeSerLysCysAsnGlnLeuLysSerProGlySerProGlyProGly 700
2051 TAAGCAAGTGCMAATGAGCTGAAGTACCTGGAGAGCCCTGGACCTGAG 2100
701 LeuProIleGluThrAlaLeuAspAspArgLysArgGlyLeuSerHisSe 717
2101 CTGCCCATTTGAACAGCGCTTGGATGATAGAGAACGAAATTTCCCATTC 2150
717 rLeuTyrSerGlyIleGluGlyLeuAspGlyLeuSerProSerArgAsnAla 734
2151 CCTCTACAGTGGGATTTGAGGGCTTGATCAATCCCCAGACGAATAGCTG 2200
734 1aLeuSerArgIleMetGlyLysSerGlyLysTyrGlnLeuSerProThr 750
2201 CCTCTACAGTGGATTAATG.....GGTAAATACCACTGCTCCCTACCA 2241
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
2242 GTGAATATGCCCCAAGATGACCTCTCATTTAAGAAATGACAGGTGGC 2291
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
2292 AGTCTCTCTCCATCTCTCTGACACAGTCTCTTCAGCTCCCATGATG 2341
784 sPValGlyPheValThrAlaAspAlaGlyThrThrAlaLysAlaIle 800
2342 ATGTGGGTTTGTACGCGAGATGCTGTGAGTGGGCCAAGGCTCCCAATC 2391
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
2392 AGTGATTCAGCCGACGCTCTTGTGATCCAGATGTTGATCCAGTCTTGC 2441
817 apHeGlnArgGluGlyPheGlyArgGlnSerMetSerGlnLysArgThrL 834
2442 TTTTCAACAGAGAATTTGGACGTCAAGATATGCAAAAAACGACAA 2491
834 ySGLnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
2492 ACCAATTTTCAGATGCCAGTCAATTGGATTTCTTAAACACAGAAATACA 2541
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
2542 AAAAGCATGATTTAGTATAGCTGACGAGACTAAACTCAATACAGTGA 2591
867 PaSPGlnLysAlaGlySerProSerArgAspAlaGlyProSerLeuGlyL 884
2592 TGACCAAAAGACAGGTTCTCCAGCGAGATGTGGTCTTCCCTGGGCTC 2641
884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
2642 TGAAGAAGTCAAGCTGTGTGAGAGCTGCAAGACCGCAGTGGCCGAGAG 2691
901 ThrLeuAsnGlyAspLysProPheHisArgProArgProArgIleIleAr 917
2692 ACTTTGAAGGGGATATCTCTTCATCGTCCAGGCGCGGATATATCAG 2741
917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaLysLysSerTyrA 934
|||||
2742 AGCGAGGGATGCAATGAGAGCTTCAGAGCTGCAATGCAAAATCTTATG 2791
934 sPlysProAlaValAspAspAspGlnGlyMetGluThrLeuGlnLeu 950
2792 ATAAACCCCGGTAGATGATGATGATGAGGATGAGACCTTGGAGAA 2841
951 AsPThrGlnLeuSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
2842 GACACAGAAAGAAATTCAGATCAGGAGAGAGTCTGATCCACAGGACG 2891
967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGlu 984
2892 TGATCAGCCCTTCCCTCTCTGGAGAGCAATTAATGAAGAACCAAGACA 2941
984 ySGLYAspLysThrAspArgLysLysAspLysThrGlyLysGlyLysLys 1000
2942 AAGGTATAGACTGATATGAAAAAAGATTAACCTGAAAAAAGAAAGAG 2991
1001 LysAspArgAspLysGlyLysAspLysMetLysAlaLysLysGlyMetLe 1017
2992 AAAGTATAGATTAAGAGAGATTAATTAATTAAGCAACAGAGGAATGCT 1034
1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034
3042 GAAGGCTTGGAGACATGTTCAAGCTTGGCAAAACATGCAAGAAATGACA 1034
1034 ySILeGlyLysThrGlyLysLysLysIleGlnGlnLysSerPheThrSer 1050
3092 AGATTAGAAAGAGGATTAATTAATTAATTAATTAATTAATTAATTAAT 1051
1051 GluGluArgLysArgMetLysGlnGlnGlnGlnGlnGlnGlnGlnGln 1067
3142 GAGGAGAGGATACGAATAGACAGACAGAGAGAGATTCACCAAAAC 3191
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIleG 1084
3192 TCGACAATTTAGGAGACGACAGCAAGCTCGACAGCGTACTGCTGAATTC 1084
1084 1nAspPheHisArgThrPheGlyCysAspAspGlnLeuMetLysGlyL 1100
3242 AAGATTTTCATCGACATTTGGCTGTGATGATGATGATGATGATGATG 3291
1101 ValSerSerTyrGlnGlySerMetAlaLeuAsnAlaArgProGlnSerP 1117
3292 GTTCTCTTATGAAAGTTCATGCTCTCTCAACGCTACAGCTCAGACCC 3341
1117 oArgGluGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysPro 1134
3342 ACCAGAAAGGCAATATGATGATGCTTGTATGCCCAAGTCAAAAGACCC 3391
1134 rGlnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
3392 GGAATTTCCAAACCCCTCACTTACACAGTAAACAGATCAACTCTGACAT 3441
1151 HisAspArgGlyLeuArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
3442 CATGATCGATACAGCTCTGAGGCAAGATTTCAACCAACCAACCAAGA 3491
1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnInProTrp 1184
3492 TGAAGATGTGAAGATGTCGCGGACGACTATAGTTTGAACAACCTGGCC 3541
1184 rAsnAlaArgProAlaThrAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
3542 CGAAGCAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
1201 ValGlnMetGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1217
3592 GTGCAATGACAGCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1217
1217 agLnaArgLysSerSerLeuProArgGlnSerArgLysAsnAlaSer 1234
3642 CCAGGCGCAGTACAGCTCTCTCGCGCAAAACAGCAAAATATGCCAGCT 3691

```


134 IansMetProLeuHisValArgArgSerSerProAlaLeuIleGly 150
688 CAAATAATGCCCTTCATGTCGACGACGATGACCCAGCTCAATGGC 151
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
738 CTCCTCACTTCTGTCAAGATAGTAATTTTCTCTGAAGAGCCCTTCAG 167
167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGluAsnT 184
788 GAAAATTCACACAGCGCTGTGACACACACCTGGCTCTCAAGAGAAC 837
184 hralaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
838 CTCCTGGAGTCTTAACCTCGACAGAGAAAGATGAACACTACAGA 887
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
888 AGCCTCCGCGGAGATACGTAAGTACTGCTTACCAATTTGAGAGACAA 937
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234
938 TCTGTGCTGCTCTGTGAGTGCACCCCAATGGTGGCAAGTGGCTGG 987
234 IuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
988 AGAAACAAAGAACAGATGAGGATGGACAGAAAGAGATACAGTCTGTT 1037
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
1038 GAACTGTGGACATCTGCACACGGGTTGGAGCATATACCAACTTTC 1087
267 rIleuAspAspMetValLysLeuValGluValProAsnAspGlyLysPro 284
1088 TCTGGATGATATGTAAGCTCGTGAAGTCCCAAGATGAGGCGCTTC 1137
284 euGlyTleHisValValProPheSerAlaArgLysGlyArgThrLeuGly 300
1138 TGGGAATCCATCTAGTGTCTTACGTCTGACGGCGCAAGCCCTGGGG 1187
301 LeuLeuValLysValArgLeuGluLysGlyLysAlaGlnHisGluAsnLe 317
1188 TTATTATGTAACGATTTGGAGAAAGTGTAAAGCTGAACATGAAATCT 1237
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
1238 TTTTCGTGAGAAATGATTCATTTGTCAAGATTATGATGGACCTTCGA 1287
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1288 ATGAGAAATTTGACACACACACATATGTTGCCCAAGCCATGCTTACA 1337
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrG 367
1338 CCCATCATTTGTTCATGTGTTCCTCGACGCAAAATAAAGACAGATAGA 1387
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
1388 ACAACTATCCCAAAGTGAAGAAACAAATTAATCAAGCCGTTTAAACC 1437
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1438 CTGACAGCCAGTATATGACACAGAGAGTGAACATGCAAGGCTTCAC 1487
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1488 ACGGTGACAGAGACCCGACCTACACCCCTGAGAGATAGACAC 1537
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1538 TCACCTCAAGACTACCTCATAGCCACACCCCTGGGAAACCAACCATCCG 1587

434 IAProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1588 CTCACAGCCCTCGCACCTCAAGAAATGATTTAGTACGACTGTATGACAGTCT 1637
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 1637
1638 TATTAACACCAAAAAATATAGCAAGAGGCTTATATATCCAGCTTAAGAAAG 1687
467 YThrGluGlyLeuGlyPheSerIlePheSerArgAspValThrIleGlyG 484
1688 TACAGAGGCTTGGATTTCAGCATCCTCCAGAGATGTAAACAATAGGTG 1737
484 LysSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1738 GCTACCTCCAAATCATGTGAAAAACATCTCCCGGGGGCGCCCATTT 1787
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
1788 CAGGATGGCCGACTTAAGCAGAGACAGACTTAAGAGGTAAATGAGACT 1837
517 IAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534
1838 AGATTATGTGGCAAAATCCCAAGAGAACTTGTTCGCTGTGGAAAGCA 1887
534 hrlYsMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
1888 CCAAGATGGAGAACTGTGAGCTTCGTGCTTTCGCGAGAAAGCGCC 1937
551 PheHisProArgLysLeuAsnAlaGluProSerGlnMetGlnIleProly 567
1938 TTCACCCCAAGGAACTGAATGCAGAGCCAAAGCAATGCAGATTCCAAA 1987
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
1988 AGAAACGAACACACAGATAGATATTTCTTACACTGTATGGCACCA 2037
584 rGgluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
2038 GGGATTTCTGACATTTGAAAGTCCCACTTAAATGATTCAGATCTGCAGCG 2087
601 LeuGlyValSerValLysGlyLysAsnArgSerLysGluAsnHisAlaAspLe 617
2088 CTGTGGTGTCAAGTGTCAAGGTTACCGGTCAAAAGAACACCAACCAATTT 2137
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
2138 GGGATTTCTGCAAGTCCATTTATATGAGAGACAGCATCTAAAGATG 2187
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 650
2188 GAAAGCTTTCGGGTGATGATCAACTGATAGCAATGAGTAATGCAATTCCTG 2237
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
2238 TTGGGCAAGACAAACCAAGATGCCATGGAACCCCTAAGAGCTTATGTC 2287
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
2288 TACTGAAGGCAATTAACGAGCAATGATCCACTTATTTGCAAGAGAGAA 2337
684 IeserLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
2338 TAAAGCAAGTCAATGAGCTGAAGTCACTGTGGAGCCCTCGAGCTGAG 2387
701 LeuProIleGluThrAlaLeuAspAspArgGluArgIleSerHisSe 717
2388 CTGCCCATTTGAACACGCTTGTGATAGAGAAACAGAAATTTCCCATTC 2437
717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
2438 CTTCTACAGTGGATTTGAGGGCTTGATGAAATCGCCCAAGCAAAATCTCG 2487
734 IaleuSerArgIleMetGlyLysGlyLysTyrGlnLeuSerProThr 750

2488 CCCCAGATGATGATG.....GGTAAATACAGACTGCTCCCTCA 2528
 751 ValasmetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
 2529 GTGAATATGCCCAAGATGACACTGCTCTTTAGAAATGACAGGTTGCC 2578
 767 OvalleuProProHISLeuSerAspGlnSerSerSerSerHisAspA 784
 2579 AGTCTTCTCCACATCTCTGACACAGTCTCTTCCAGCTCCCATGATG 2628
 784 sPvalIglyPheValThrAlaAspAlaGlyThrTPalaIysAlaIle 800
 2629 ATGGGGGTTGTGACGGCAGATGCTGTTACTGGCCAGGCTGCATTC 2678
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValIleAl 817
 2679 AGTATTCAGCCGACTGCTCTTTGAGTCCAGATGTGATCCAGTTCTTGC 2728
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
 2729 TTTTCAAGAGAGATTTGGACGTCAGAGTATGTCAGAAAAAGSCACAA 2778
 834 ysgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
 2779 AGCAATTTTCAGATGCCAGTCAATTGATTTGTTAAAAACAGAAAAATCA 2828
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
 2829 AAAAGCATGATTTAGTTATAGTGCAGACGACTAAACTCAATACAGTGA 2878
 867 PaSPglNLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 2879 TGACACAGAAACAGCTCTCCACAGACAGATGTGGGTCCTCCCGGTC 2928
 884 eulysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
 2929 TGAAGAACTCAAGCTCATTTGAGAGTGTGACAGCCGAGTTGCCGAGTGT 2978
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleAr 917
 2979 ACTTTGAATGGGATATTTCTTTCCATCTCCACAGCCCGCGATATATAG 3028
 917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyra 934
 3029 AGCGAGGGAGATGCAATGAGACTTCAGAGCTGCCATCGACAAACTTATG 3078
 934 sPlusProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950
 3079 ATMAACCCGCGGTAGATGATGATGAAAGCATGAGACCTTGGAMAGA 3128
 951 AspThrGlnGlnSerSerArgSerGlyArgGlnSerValSerThrAlaSe 967
 3129 GACACAGAAAGAAAGTTCAAGATCAGGAGAGAGAGCTGTATCCACAGCCAG 3178
 967 rAspGlnProSerHisSerLeuGlnIuArgGlnMetAsnGlyAsnGlnGlu 984
 3179 TGATCAGGCTTCCACCTCTCTGAGAGACAAATGAAATGAAACCAAGAGA 3228
 984 ysgIlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
 3229 AAGGTGATAAACTGATAGAAAAAAGATMAACTGGAAGAAAAAGAAAG 3278
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
 3279 AAAATATGAGATTAAGAGAAAGATTAATGAAGCAAGAAAGGAAATCT 3328
 1017 uLysGlyLeuGlnLysAspMetPheArgPheGlyLysHisArgLysAspAl 1034
 3329 GAAGGGCTTGGAGACATGTTGAGTTTGGCAAAACATCGAAAAAGATGACA 3378
 1034 ysiIleGluLysThrGlyLysIleLysIleGlnGlnSerPheThrSerL 1050

3379 AGATTGAGAAAAAGGTAAATATAAATACAGAACTCTTACATCAGAA 3428
 1051 GluGluArgGlyLeuArgMetLysGlnGluGlnGluArgGlyIleGlnAlaLysTh 1067
 3429 GAGGAGAGATTCGAATGATACAGCAGCAGCAGAGAGATTTCAAGCCAAAC 3478
 1067 rArgGluPheArgGluArgGlnAlaArgGluIuArgAspTyraIleGluIleG 1084
 3479 TCGAAGAAATTTAGGAAGACAGCAAGCTCGAGAGCGTACATAGCTGAATTC 3528
 1084 lAspPheHisArgThrPheGlyCysAspAspGluLeuMetGlyGlyGly 1100
 3529 AAGATTTTCATCGGACATTTGGCTGTGATGATGATTAATATATGGGGA 3578
 1101 ValSerSerTyrgLugLysSerMetAlaLeuAsnAlaArgProGlnSerPr 1117
 3579 GTTCTTCTTATGAAGTTCCTCATGCTCTCAACGCTACAGCTCAGAGCCC 3628
 1117 oArgGluGlyHisMetAspAlaLeuTyraIleGlnValLysLysProA 1134
 3629 ACGAAGAGGCGATATGATGATGATGCTTTGTATGCCCAAGTCAAGAGCCGC 3678
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
 3679 GGAATTCAAAACCTCACCTGTAGACAGTAAACAGATCAACTCTTAGCAAT 3728
 1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
 3729 CATGATCGGATACAGCTCTGAGGCAAGATTTCAAGCAAGCAAGCAGACA 3778
 1167 pGluAspValGluAspArgArgArgThrTyserPheGlnGlnProTrp 1184
 3779 TGAAGATGAGAAATCGTCGGCGGACCTATAGTTTGAACAACCTGGCC 3828
 1184 rOAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
 3829 CGAAGCGACGGCGGCGAGCGAGCGGCGGACCTCGGTGTCCTGGAG 3878
 1201 ValGlnMetGlnArgGlnArgGlnGluIuArgGlnSerSerGlnGlnAl 1217
 3879 GTGCGATGTCAGCGCGCAGCGCCAGCAGAGAGCGCGAGAGCTCCACAGCGC 3928
 1217 aGlnArgGlnTyserSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234
 3929 CCAGCGCCAGTACAGCTCTCTCGCGCAAGCAGAAATAATGCGAGCT 3978
 1234 ervaSerGlnAspSerTrpGluGlnAsnTyserProGlyLugLysPhe 1250
 3979 CGGTCTCCAGGACTCTGGGAGAGAACTACTCCCTGGGGAGAGGCTTC 4028
 1251 GlnSerAlaLysGluAsnProArgTyserSerTyrgLingLysSerArgAs 1267
 4029 CAGAGTCCCAAGAGAACCCCGAGTACTCCAGTTCACAAAGCTCCAGGAA 4078
 1267 nGlyTyrlLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThrG 1284
 4079 CGGCTACTCTGGAGAGACATGCTTCAACGCCAGGGCATGCTGGAACATC 4128
 1284 lngluLeuLeuArgGlnGluGlnArgArgLysGluGlnGlnMetLysLys 1300
 4129 AGGACTCTCTTCCGCAAGAAACAGAGCGAAGACAGCAGCATGTGAAGAAG 4178
 1301 GlnProProSerGluGlyProSerAsnTyraSpSerTyrgLysLysValG 1317
 4179 CAGCTCTCTCCGAGAGGCGCCAGCAACTATGACTGTATAGAAAGTCCA 4228
 1317 nAspProSerTyraIaProProLysGlyProPheArgGlnAspValProp 1334
 4229 GGACCCCAAGTTACGCCCTCCCAAGGGGCCCTTCGGCAAGATGTGCCCC 4278
 1334 roSerProSerGlnValAlaArgLeuAsnArgGlnGlnThrProGluLys 1350
 4279 CTTCCCTTCTCAGGTTGAGAGCTGAACAGACTTCAGACTCTCGAANA 4328

```

1351 G1YARpProPheTySer 1356
|||||
4329 GGGAGGCCCTTCATATCC 4346
Seq_name: 9b_pr:AF467004

seq_documentation block:
LOCUS AF467004 3960 bp mRNA linear PRI 07-FEB-2002
DEFINITION Homo sapiens partitioning-defective 3 protein splice variant d
ACCESSION AF467004
VERSION AF467004.1 GI:18568347
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3960)
AUTHORS Gao,L., Macara,I.G. and Joberty,G.
TITLE Multiple splice variants of PAR3 and of a novel related gene,
JOURNAL PAR3L, produce functionally different proteins
2 (bases 1 to 3960)
AUTHORS Gao,L., Macara,I.G. and Joberty,G.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2002) Center for Cell Signaling, University of
Virginia, PO Box 800577, Charlottesville, VA 22908, USA
FEATURES
source
1..3960
location/Qualifiers
gene
1..3960
db_xref="taxon:9606"
/gene="Homo sapiens"
1..3960
/gene="PAR3"
1..3960
/note="PAR3"
1..3960
/note="PAR3d, multi-PDZ domain containing protein;
alternatively spliced"
/codon_start=1
/product="partitioning-defective 3 protein splice variant
d"
/protein_id="AAL76044.1"
/db_xref="GI:18568348"
/translation="MKYVGRGRVYVPCGQGHMKVPSLIQOATRYRKAIDPNY
IIOVRLHGGDGLDLDITCDVADKDRILAVDEDDPHGGDGSASSTGQSPF
ISSEPSRKNPTREKSTTAGETKONTAGSPKCKKDEYRSLDITGLSTVDSNF
RSLSSASHPWGKWLKQKQDEDEEDNSRYEPKSHADTLEHIDPSTMSNDQKQNA
VPNDGGELGTHVYFSARGRTGLLVRLKRGKAEHENTFRENQIVINQDGLNR
RFPKQDHPKQKMPRIIMFHVPAANKRQKPOLSKSEKNNYSSSFSDSYDNR
VNSAGLHVQAPRLNHPPEQIDSHSLPSSASPKSPSAPASAPQVVFSTYSG
LYNKTKGRLNQLKKGEGLEPSTSDYITGSAPIVYKNIIPGALIDGRKAG
DRLIEVNGVDLGKQSEVYSLRSTKEGTVSLVPRQDAHPRELAIDGRKAG
KETKADEDEDIVLPDGTREFLFEVPIINDSGSAGLVSVKAKNMLKGIQMFRTQAT
INGASAKDGRILVNDOLIVNGESLGNQAMETLRRSNKRENNHNDLCTFKVST
RTSKCNELKSPSPGPELPIETRLDREKRIISHSYSGIEGLDSSPSNNALSRITG
ESGKYOLSPVNPMDPVIIEEDRLPVLPEPHLSDQSSSHDDYGFYTAAGWMA
AIDSDADSLSPVDVPIATQKRGFGQSKSEKRTQESDASLDVFTAKSGMADG
IIRGCGNESPRALDKSYDKPAVDDDEGMELEEDPESSRGRESVSTASQPSH
SLEKONGNOKERAKDKTKKDKTKGKKDKRKKKAKKAKMLKGIQMFRTQAT
REFEROAREDEDAEIODPRTGCDDELMGCVSSTGSAKAKMLKGIQMFRTQAT
LYAVKTRKSPSPVDSNKTSPSNHRIORLROFQDAKODEDEVERRRATYFQPD
PNAPRATQSGRHSVVEVOMOROEERESSQAKORQSSLPKRSRRNASSVDQWE
KQSPSEGFQSAENRYSSYGSRNGVLGGHGNARMLMETOLERQDREKQDMK
KQSPSEPSNYTSYKRYQDPSTAPPKGPFQDVPSPSPQVARNRLQTEKGRPTYS"
814..1074
/gene="PAR3"
/note="Region: PDZ domain"
1378..1638
/gene="PAR3"
/note="Region: PDZ domain"
misc_feature
1378..1638
/note="Region: PDZ domain"

```

```

misc_feature 1765..2028
/gene="PAR3"
/note="Region: PDZ domain"
BASE COUNT 1170 a 910 c 1048 g 832 t
ORIGIN

alignment_scores:
Quality: 1025.00 Length: 1025
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AF467004
Align seg 1/1 to: AF467004 from: 1 to: 3960
1 MetLysValThrValCysPheGlyArgThrArgValValValProCysG1 17
|||||
1 ATGAAAGTACCGCTGCTCTCGACGCGACCGGCGTGGTGGCGCGG 50
17 YASpGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
51 GGAGCGGCACATGAAGATTTCAGCCTCATCCACGAGCGGTGACCGCGT 100
34 YIRArgYsaAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
101 ACCGGAGGCGCATCCGCAAGATCCAACTGATGATACAGTGCATGCC 150
51 LeuGlnHisGlyYASpGlyIleLeuAspLeuAspAspIleLeuGlySAs 67
151 TTGGAGACATGGAGATGAGGATGATAGACCTTGATGATGATCTTTGTA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnGlnAsp 84
201 TGTAGACAGCATTAAGACAGACTGGTAGCAGTGTATATGAGAGATC 250
84 roHisHisGlyGlyYASpGlyThrSerAlaSerThrGlyThrGlnSer 100
251 CACATTCACGAGGTGATGCGCACAGTCCAGTCCACGCGTACCAAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
301 CCGAGATATTTGGTAGAGCTTGACCGACCAACATGTCTCACCTTTCA 350
117 nProYrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAAGCACAGTGAATTTGAGTCCACCTTAGTCTTCGAG 400
134 IaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleG1 150
401 CAATATGCTCTTCATGATTCGACGCGATGAGTACCGCTCTAATTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
451 CTCGCCCTCTCTGATGATATATTTCTCTCTGAAGCCTTCAG 500
167 gLysAsnProIleArgTyrPserThrThrAlaGlyPheLeuLysGlnAspT 184
501 GAATAATCCACACGCTGTGTCAACACAGCTGGCTTCTTCAGCAACA 550
184 hValAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTGCTGGAGAGTCTTAAACCGCGACACGAAAGAACTAAACATACAGA 600
201 SerLeuProArgAspThrSerAsnTyrPserAsnGlnPheGlnArgAspAs 217
601 AGCTCCCGCGGAGTACTATGTAAGTGTATACCAATTTCAGAGAGCA 650
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
651 TGCTCGTCTGCTCTGAGTGCAGTCAACCCCAATGCTGGCAAGTGGCTGG 700

```

```

234 IuysgIngluGlnAspGluAspGlyThrGluGlnAspAsnSerArgVal 250
|||||
701 AGAAACAGAGACAGATGGAGATGGACAGAAAGAGATTAACAGTCGCTT 750
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
|||||
751 GAACCTTTGGACATGCTGACACGGGTTTGGAGCATATACCCAACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyProL 284
|||||
801 TCTGATGATATGTGTAAGCTCTAGAAGTCCCAACGATGGAGGCTC 850
284 euGlyIleHisValValProPheSerAlaArgGlyValArgThrLeuGly 300
|||||
851 TGGGAATCCATGTAGTCCCTTTCAGTCTCGAGCGGCGAGAACCCCTGGG 900
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGlnHisGluAsnLe 317
|||||
901 TTATTTAGTAAACGATGGAGAAAGGTGGTAAAGCTGAAACTGAATAACT 950
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
|||||
951 TTTTCGTGAGATATGATGCAATGTTCAGAGATTAAATGATGGGACCTTGA 1000
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
|||||
1001 ATAGAAATTTGAAACAGCAACACATATGTTTCCCAAGCCATCGTACA 1050
351 ProIleIleThrPheHisValValProAlaAlaAsnLysGluGlnArg 367
|||||
1051 CCCATCTTTGGTTCATGTGTTCTCTGCAACAAATTAAGACGATATGA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrrTyrSerSerArgPheSer 384
|||||
1101 ACACTATCCCAAGTGAGAAAGCAATTAATTAATCAAGCCGTTTAGCC 1150
384 roAspSerGlnTyrrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
|||||
1151 CTGACACCCAGTATATTTGACAAACAGAGTGTGAACAGTGCAGGCTTCAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
|||||
1201 ACGGTGAGAGAGACACCCCGACTGAACACCCCGCTGACGAGATAGACTC 1250
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
|||||
1251 TCACCTCAAGACTACCTCATAGCGCACACCCCTGGGAAAAACACATCCG 1300
434 lAspAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
|||||
1301 CTCGAGCCCTGGCACCTCAGAAATGATTTAGTACGACTGTAAAGCAGTGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGly 467
|||||
1351 TATTAACACCAAAAATAATGAGCAAGAGCTTAATATCCAGCTTAAGAAAGG 1400
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
|||||
1401 TAGAAGAGGTTTGGATTACACATCAGCTTCCAGAGATGTAAACATAGTG 1450
484 lYserAlaProIleTyrrValLysAsnIleLeuProArgGlyAlaAlaIle 500
|||||
1451 GCTCAGCTCCCAATCTATGTGAAAAAATTTCCCGGGGGGGCGGCATTC 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517
|||||
1501 CAGATGCGCCAGCTTAAGGCAAGAGACGACTTAAAGGTAAATGAGAGT 1550
517 lAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
|||||
1551 AGATTTATGTGGCAATCCCAAGAGGAAGTGTGCTGTTGGAACAACA 1600
534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGlnAspAla 550
|||||
1601 CCAAGATGGAAGGAAGCTGTGACCTTCGTGCTTTCGCCAGAGAAACCC 1650
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
|||||
1651 TTCCACCCCAAGGGAAGTAAATCGACAGCAACCCAGATGCAAGATTCCAA 1700
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
|||||
1701 AGAAACGAAACAGAGATGAGATATGTCTTAATACACTGATGGACCA 1750
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
|||||
1751 GGAATTTCTGACATTTGTAAGTCCACCTTAATGATTCAGAGATCTCCACGC 1800
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
|||||
1801 CTTGGTGTCACTGTCAAGATGACCGGTCAAAAGAAACCAAGCAGATTT 1850
617 uGlyIlePheValLysSerIleIleAsnGlyValAlaAlaSerLysAspG 634
|||||
1851 GGGATCTTTGTCAAGTCCATTAATTAATGAGAGGACACATCTAAAGATG 1900
634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
|||||
1901 GAAGCTTCGGGTGATGATCACTGATAGCAGTAAATGGAGAAATCCCTG 1950
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrIleuArgArgSerMetSe 667
|||||
1951 TTGGGCAAGACAAACCAAGATGCCATGGAACCCCTAAAGAGTCTATATTC 2000
667 rThrGluGlnLysAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
|||||
2001 TACTGAAGCAATAAAGAGAGATGATCAGCTTATGTTCGAAGAGAGA 2050
684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
|||||
2051 TAAAGAACTGCAATGAGCTGAAGTCACTGGAGGACCCCTGAGACTAG 2100
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
|||||
2101 CTGCCCATTTGAACAGCGCTTGATGATATGAGACAAACAAATTTCCCATTC 2150
717 rLeuTySerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
|||||
2151 CCTCAGAGTGGGATTTGAGGGCTTGATGAATCCCGCAGCAAGAAATGCTG 2200
734 lAluSerArgIleMetGlyGluSerGlyLysTyrrGlnLeuSerProThr 750
|||||
2201 CCTCAGTAGGATATATGGGTGAGTCAGGTAAATATACACCTGCCCCCTACA 2250
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
|||||
2251 GTGAATATGCCCCAAGATGACCTGTCAATTAATGAGAAATGAGAGGTTGCC 2300
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
|||||
2301 AGTGTTCTTCACATCTCTCTGACCAATCCCTCCAGCTCCCATGATG 2350
784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800
|||||
2351 ATGTGGGTTTGTGACGCGCAGATCTGTACTGTGGCCCAAGGCTCAATTC 2400
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
|||||
2401 AGTGATTCAGCCGACTGCTTTTGAGTCCAGATGTTGATCCAGTTCCTTGC 2450
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGlnLysArgThrL 834
|||||
2451 TTTTCAAGAGAGATTTTGGACGTCAAGATATGTCAGAAAAAAGCACACA 2500
834 yGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
|||||

```

2501 AGCAATTTTCAGATGCCAGCTCAATTCGATTAAACAGCAAAATCA 2550
 851 LysSerMetAspLeuGlyTlleAlaAspGluThrLysLeuAsnThrValAs 867
 2551 AAAAGCATGATTTAGTATACCTGACGAGACTAAACCAATACAGTGGA 2600
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 2601 TCACCAAGAAAGCAGGTCTCCACAGAGATGTGGTCCCTCCCTGGGTC 2650
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGlyVal 900
 2651 TGAAGAACTCAAGCTCGTTGACAGAGCTCGACAGCCGACGTTGCCAGAG 2700
 901 ThrLeuAsnGlyAspGlnLeuProPheHisArgProArgProArgTlleLeu 917
 2701 ACTTTGATGGGATATTCCTTCCATCGTCGCAAGCGCGGATTAATCAG 2750
 917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrA 934
 2751 AGCGAGGGGATGCAATGAGAGCTTCAGAGCTGCCATCGCAAAATCTTAG 2800
 934 sPLysProAlaValAlaAspAspAspGlnGlyMetGlnThrLeuGlnGlu 950
 2801 ATAAACCCCGGTAGATGATGATGAGAGCGATGACACCTTGGAAGA 2850
 951 AspThrGlnGlnSerSerArgSerGlyArgGlnSerValSerThrAlaSe 967
 2851 GACACAGAAAGAAATTCAGATCAGAGGAGAGAGTGTGTATCCACAGCCG 2900
 967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlnGlnGlnGlu 984
 2901 TGATCGACCTTCCACTCTCTGAGAGCAAAATGAAATGAAACCAAGA 2950
 984 yGlnLysLysThrAspArgLysLysAspLysThrGlyLysGlnLysLys 1000
 2951 AAGGTCATAGACTGATAGAAAAAGATAAACGCGAAAAAGAAAGAG 3000
 1001 LysAspArgAspLysGlnLysAspLysMetLysAlaLysGlnMetLe 1017
 3001 AAGATAGAGATAGAGAGAGAGATAAATGAAAGCAAGAGGAGGAAATGCT 3050
 1017 uLysGlnLeuGlyAspMetPheArg 1025
 3051 GAAGGCTTGGAGACATGTCAGG 3075
 seq_name: gb_pr:AF196186
 seq_documentation_block:
 LOCUS AF196186 3470 bp mRNA linear PRI 16-NOV-2001
 DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein
 ACCESSION AF196186
 VERSION AF196186.1 GI:13491611
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS 1 (bases 1 to 3470)
 TITLE Fang, C.M. and Xu, Y. H.
 JOURNAL Down-regulated expression of atypical PKC-binding domain deleted
 MEDLINE Cell Res. 11 (3), 223-229 (2001)
 PUBMED 21520080
 JOURNAL 11642408
 REFERENCE 2 (bases 1 to 3470)
 AUTHORS Fang, C. and Xu, Y.
 TITLE Exon/Intron Structure and Splicing Variants of a Novel Human
 JOURNAL Polarity Gene, hASP
 REFERENCE 3 (bases 1 to 3470)
 AUTHORS Fang, C. and Xu, Y.

TITLE Direct Submission
 JOURNAL Submitted (18-OCT-1999) Laboratory of Molecular and Cellular
 Oncology, Shanghai Institutes for Biological Sciences, Chinese
 Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P.R. China
 FEATURES
 source
 1..3470
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /map="10p11.2"
 288..3383
 /note="hasIPsa"
 /codon_start=1
 /product="atypical PKC isotype-specific interacting
 protein short variant"
 /protein_id="AAK27892.1"
 /db_xref="GI:13491612"
 /translation="MKTVVCFGRTRVVVPCGGMKVFSLIQAVTRRAIAKDPNY
 MIOVRLRHGDGGLDDLDLIDVADDRILVAVPFDQDPHHGSDTSASTQSPF
 IFGSLGTNNVSARQPVQATSEIETVPSVLAMPPLHVRSSDPALIGLSTQSPF
 SSEPSRKNPTRMSTTAGFLKQNTAGSPKCTDRKDNYSLSLPRDSMNSQPDONA
 VPNDGSLGIHVVPFSARGRTGLLVKRLKGGKAHEMLFRENQIVINDGLRN
 RRFDOAHMFQAMRTPPIWFHVVPANPKQDYDOLDSQENNNYSSRFPSDSTIDNR
 SVNSAGLTVOARAPRLNHPPEQIDSHSLRLPSHAPSGKPSAPASAPQNVSTTVSSG
 YNTRKIGRLNIOLEKTEGEGFITSROYTGGSAP IYKNIIPGAAIIOBRKAG
 DRLIEVGVLDVKGSEEVSLRSTMBETVSLVFRDQARHRELNAESOMIPI
 KETRAEDEDIYLPDGRREPLTREVPLINDSGSAGLVSVGNRSKRNHADIIFYVSI
 INGGAASKDRRLRNDDLIIVNBSGLKTNQDAMETLVSVGNRSKRNHADIIFYVSI
 RISCHNELKSGSPGPELPIETALDREERRISHSYSGIEGLDESPNNAALSRMG
 KYQLSPTVNMQDDVTIIEDDRLLVPLPRLHSDSSSSHDQGFVADAGCAAAIS
 DSADCSLSPDVPVLAFOREGFRQSGSEKRTFQFADASOLDFVTRKSKMDGLAD
 ETKLNTVDQKASPSRDVSPSGILKSSLSIESIOTFAVAYTLNGLDIPHRPPIIR
 GRGNESEFRAALIKSYDKPAVDNDDSGMETLEEDTESSRSRESVSTASDQPSHLE
 R" ROMNGNEKEDKTRDKKTKGKKRDKREKDKMAKGMKLGIDMFSIAKLKPEK
 BASE COUNT 1003 a 811 c 934 g 721 t 1 others
 ORIGIN
 alignment_scores:
 Quality: 781.00 Length: 1024
 Ratio: 0.765 Gaps: 1
 Percent Similarity: 99.707 Percent Identity: 99.707
 alignment_block:
 US-09-757-781-2 x AF196186 ..
 Align seg 1/1 to: AF196186 from: 1 to: 3470
 1 MetLysValThrValCysPheGlyArgThrArgValValValProCysG 17
 288 ATGAAGTGAAGCCGTCTCGGAGGAGCCGGGTGCTGCGCGCGG 337
 17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
 338 GGAGCGGCACATGAAAGTTTTCAGCTCATCCAGCAGCGGGTGAACCCGCT 387
 34 yArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
 388 ACCGGAAGGCCATCCCAAGAGATCAAACTACGTAGATACAGCTGATCCG 437
 51 LeuGlnHisGlyAspGlyLysLeuAspLeuAspAspIleLeuCysAs 67
 438 TTGGAACATGAGATGAGAGAACTAGACCTTGATGACATTCCTTGTGA 487
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnGlnAsp 84
 488 TGTAGCAGACGATTAAGACAGACTGTGATGATGACGAGATC 537
 84 roHisHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
 538 CACATCAGCAGAGTGATGACACCACTGCTCACGGGTACCCAGAGC 587

101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
 |||||
 588 CAGAGATATTGGTAGTGAGCTGGCACCAACAATGCTCAGCCTTCA 637
 |||||
 117 nProTyGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
 |||||
 638 GCCTTACCAAGCAACAGTGAATTAAGGTACACCTTCAGTCCCTCGAC 687
 |||||
 134 IaasMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
 |||||
 688 CAAATATAGCCTTTCATGTTGACGACGAGTACGCCACGCTCTAATTGCG 737
 |||||
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
 |||||
 738 CTCCTCACTCTGTGAGTAGTAATTTTCTCTGAAGAGCCTTCAG 787
 |||||
 167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
 |||||
 788 GAAAAATCCACACGCTGGTCACACACGCTGCTCCCAAGCAGAAC 837
 |||||
 184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyArg 200
 |||||
 838 CTGCTGGAGAGCTTAAAACTGCGACAGAAAGATGAATACTACAGA 887
 |||||
 201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
 |||||
 888 AGCCTCCCGCGGATACTAGTAAGTCTTCAACCAATTTCAGAGAACAA 937
 |||||
 217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234
 |||||
 938 TGCCTGCTGCTCTGTGAGTGCACATCCCAATGTGGGCAAGTGGCGG 987
 |||||
 234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
 |||||
 988 AGAAACAGAACAGAGATGAGGATGGACAGAAAGGATTAACAGTCGCTT 1037
 |||||
 251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheSe 267
 |||||
 1038 GAACCTGTGGACATGCTGACACAGGTTTGGACATATACCAACTTTTC 1087
 |||||
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284
 |||||
 1088 TCTGTGATGATGTGTAAGCTCTGTAAGTCCCAACGATGAGAGGCTTC 1137
 |||||
 284 euGlyIleHisValAlaProPheSerAlaArgGlyLysArgThrLeuGly 300
 |||||
 1138 TGGGAATCCATGTAGTGCCTTTCAGTCTCGAGCGGCGAACCCTGGGG 1187
 |||||
 301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLe 317
 |||||
 1188 TTATTAGTAAAAAGCATGGAGAAAGTGTGAAGCTGMACTGMAAATCT 1237
 |||||
 317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyLysLeuArg 334
 |||||
 1238 TTTTCGTGAGATGATGTGCATGTTCAGGATTAAGATGCGAACCCTCGAA 1287
 |||||
 334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
 |||||
 1288 ATAGAAGATTGGAACAGCAACACATATGTTCCGCAAGCCATCGGTACA 1337
 |||||
 351 ProIleIleTrpPheHisValAlaProAlaAlaAsnLysGluGlnTyArg 367
 |||||
 1338 CCCATCATTTGTTCATGTGTCTCTGACCAATAAAGAGCAGTATAGA 1387
 |||||
 367 uGlnLeuSerGlnSerGluLysAsnAsnTyTrpSerSerArgPheSerP 384
 |||||
 1388 ACAACTATCCCAAAAGTAGAAGAAACAATTACTATTCAAGCCGTTTTACCC 1437
 |||||
 384 roAspSerGlnTyrlLeuAspAsnArgSerValAsnSerAlaGlyLeuHis 400
 |||||
 1438 CTGACAGCCATATATTGCACAACAGAGACTGTGAACAGTGCAGGCGCTTAC 1487

401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
 |||||
 1488 ACGGTGCAGAGAGACACCCCGACTGAACACCCGCGCTGAGAGATGACTC 1537
 |||||
 417 rHisSerArgLeuProHisSerSerAlaHisProSerGlyLysProProSera 434
 |||||
 1538 TCACCTCAAGACTACCTCATACGCGACACCCCTCGGGAAGAACACCATCCG 1587
 |||||
 434 lProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
 |||||
 1588 CTCACGCTCGGCACCTCAGAAATGATTTAGTACACACTGAAGAGTGGT 1637
 |||||
 451 TyraSnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG1 467
 |||||
 1638 TATAACACCAAAAAAATAGCAAGCGCTTAATATACCAAGTTAAGAAAGG 1687
 |||||
 467 yThrGluGluLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
 |||||
 1688 TACAGAAAGTTTGGATTTCAGCATCTTCCAGAGATGTAAACAATAGGTG 1737
 |||||
 484 lYserAlaProIleTyValLysAsnIleLeuProArgGlyAlaAlaIle 500
 |||||
 1738 GCTCAGCTCCATCTATGTGAAAAACATTTCCCGCGGGGGCGGCATT 1787
 |||||
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
 |||||
 1788 CAGGATGCGCCGACTTAAGCAGAGACAGACTTATAGAGGTAAATGAGAT 1837
 |||||
 517 lAspLeuValGlyLysSerGlnGluGluValSerLeuLeuArgSert 534
 |||||
 1838 AGATTATGTGGCAAAATCCCAAGAGAAAGTGTTCGCTGTGAAGAAC 1887
 |||||
 534 hTrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
 |||||
 1888 CCAAGATGGAAGAACTGTGACCTTCTGTGCTTCGCGCAGGAACACGCC 1937
 |||||
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
 |||||
 1938 TTCCACCCCAAGGAAGTGAATGCAGAGCAAGCCAGATGCAGATTCANA 1987
 |||||
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThr 584
 |||||
 1988 AGAAACGAAGCAGAGAGATGAGATATTGTCTTACACCTGATGACACA 2037
 |||||
 584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
 |||||
 2038 GGGAAATTTCTGACATTTGAAGTCCCATTAATGATTCAGAGATCTCAGGC 2087
 |||||
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
 |||||
 2088 CTTGGTGTCAAGTCAAGGTAAACCGGTCAAAAGACAACACCGCAGATT 2137
 |||||
 617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaHisSerLysAspG 634
 |||||
 2138 GGAATCTTTTGCAGTCCATTATTAATGAGAGACAGCATCTTAAGAGTG 2187
 |||||
 634 lArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysLeuSer 650
 |||||
 2188 GAAGCTTCGGGTGAATGATCAACTGATAGCAGTAAATGGAGAAATCCGCG 2237
 |||||
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
 |||||
 2238 TTGGCAAGACAAACCAAGATGCCATGGAACCCCTTAACAAAGGTCTATGTC 2287
 |||||
 667 rThrGluGluLysAsnLysArgLysMetIleGlnLeuIleValAlaArgArg 684
 |||||
 2288 TACTGAAGCAATTAACAGAGATATCCAGCTTATGTGTGCAAGAGAGAA 2337
 |||||
 684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
 |||||
 2338 TAAGCAAGTGCAGATGAGCTGAAGTCACTCGGAGGCCCTCGAGACTGAG 2387
 |||||
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717

```

2388 CTGCCCATGAAACAGCGTTGGATGATAGAGAACGAGAAATTTCCCATC 2437
717 TLeuTySerGlyIleGluIleuAspGluSerProSerArgAsnAla 754
2438 CCTCTACAGTGGGATGAGGGGCTTGATGAAATGCCCCAGAGAAATGCTG 2487
734 TAluSerArgIleMetGlyIleuSerGlyTyrgIleuSerProthr 750
2488 CCCCTAGTAAATATG.....GGTAAATACCAAGCTGTCCTCCATCA 2528
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
2529 GTGAATATGCCCCAAGATGACACTGTCATTTATGAAAGATGACAGGTGCC 2578
767 OValleuProProHleuSerAspGlnSerSerSerSerHleuAspA 784
2579 AGTGCCTTCCCTCCATCTCTGACACAGTCCCTTCCACCTCCCATGATG 2628
784 sPValGlyPheValThrAlaAspAlaGlyThrTPAlaAlaAlaIle 800
2629 ATGTGGGTTTGTGACGCGAGATGCTGTACTTGGCCCAAGCTCCTCAATC 2678
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValIleu 817
2679 AGTGAATTCAGCCGACTGCTTTGAGTCCAGATGTTGATCCAGTTCCTG 2728
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluIysArgThr 834
2729 TTTTTCACAGAGAGATTTGGACGTGACAGATGTCAGATAAAGCAACACA 2778
834 ySgIlnPheSerAspAlaSerGlnIleuAspPheValIysThrArgIys 850
2779 ACCAATTTTCAGATGCGATCAATTGATTTCTTAAACACGAAATCA 2828
851 LysSerMetAspLeuGlyIleAlaAspGluThrIysLeuAsnThrVal 867
2829 AAAACCATGAGATTAAGTATAGCTGACGACAGACTCAATCATACAGTGA 2878
867 pAspGlnIysAlaGlySerProSerArgAspValGlyProSerLeuGly 884
2879 TGACACAGAAACAGGTTCTCCACAGACAGATGCGCTCCCTGGGCTC 2928
884 eLysIysLysSerSerLeuGlnSerLeuGlnThrAlaValAlaGlyVal 900
2929 TGAAGAACTCAAGCTCATTTGGAGAGCTGCGACCGCAGTTGCCGAGCTG 2978
901 ThrLeuAsnGlyAspIleProPheHleuArgProArgProArgIleIle 917
2979 ACTTTGAATGGGATATTTCTTTCATGTCACAGCGCGGATTAATCAG 3028
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyra 934
3029 AGGACAGGAGATGCAATGACAGCTTCACAGCTGCCATCGACAAATCTTATG 3078
934 sPLysProAlaValAspAspAspGluGlyMetGluThrLeuGluGlu 950
3079 ATTAACCCCGGATGATGATGATGAAGCATGAGACTTGGAAGAA 3128
951 AspThrGluIleuSerSerArgSerGlyArgGluSerValSerThrAla 967
3129 GACACAGAAAGTTCAGATGAGGAGAGAGAGTGTGTATCCACACCCAG 3178
967 rAspGlnProSerHleuSerLeuGluArgIleMetAsnGlyAsnGlnGlu 984
3179 TGAATCAGCCCTCCACTCTCTGAGAGACAAATGATGAAACCAAGACA 3228
984 ySgIysAspLysThrAspArgLysLysAspLysThrGlyLysGluIysLys 1000
3229 AAGGTATTAAGACTGATAGAAAAAGATTAACCTGAAAAAGAAAAAGAG 3278
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysGlyMetLe 1017

```

```

3279 AAAGATACAGATTAAGACAGATTAATGAAGCCAGAGAGGAAATGCT 3328
1017 uTySgIleuGlnIysAspMetPhe 1024
3329 GAAGGCTTGGAGACATGCTC 3350
seq_name: gb-pr:AK024668
seq_documentation_block:
LOCUS AK024668 2116 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21015 f15, clone CAE05730, highly similar to
ACCESSION AF252293 Homo sapiens PAR3 (PAR3) mRNA.
VERSION AK024668.1 GI:10437004
KEYWORDS oligo capping; f15 (full insert sequence).
SOURCE Homo sapiens primary endothelial cells of human coronary artery
cDNA to mRNA, clone_11b:CAE clone:CAE05730.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (siles)
REFERENCE Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2116)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan, cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'-3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source location/Qualifiers
1. 2116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="Primary endothelial cells of human coronary
artery"
/clone="CAE05730"
/clone_11b="CAE"
/note="cloning vector PME18SFL3"
1. 2116
/note="highly similar to AF252293 Homo sapiens PAR3 (PAR3)
mRNA"
BASE COUNT 651 a 476 c 571 g 418 t
ORIGIN
alignment_scores:
Quality: 614.00 Length: 614
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AK024668
Align seg 1/1 to: AK024668 from: 1 to: 2116
743 GLyTyTyrgIleuSerProThrValAsnMetProGlnAspAspThrVal 759
|||||
200 GGTAATACCAAGCTGCTCCCTACAGTGAATATGCCCCAAGATGACACTGT 249

```

759 IIEIIEGLIuAspAspArgLeuProValIleuProHisLeuSerAspG 776
 250 CATTATGAAAGATGACAGTGTGCGAGTGTCCACATCTCTCTGACC 299
 776 InsertSerSerSerHisAspAspValGlyPheValThrAlaAspAla 792
 300 AGTCCTCTCCAGCTCCATGATGATGGGGTTGTGACGGCAGATGCT 349
 793 GlyThrTrpAlaLysAlaAlaIleSerAspSerAlaAspCysSerLeuSe 809
 350 GGTAAGTGGGCAAGGCTGCATCAGTATTCAGCCGACGCTCTTTGAG 399
 809 rProAspValAspProValIleuAlaPheGlnArgGlyGlyPheGlyArgG 826
 400 TCAGATGTTGATCCAGTCTCTGCTTTCAACGAGAGGATTTTGAGCTC 449
 826 InsertSerGluLysArgThrLysGlnPheSerAspAlaSerGlnLeu 842
 450 AGAGTATGTCAGAAAGCGCAAGCAATTTTCAGATGCCAGTCAATGTG 499
 843 AspPheValLysThrArgLysSerLysSerMetAspLeuGlyIleAlaAs 859
 500 GATTGCTTAAACACGAAATATCAAAAGCATGATTTAGTATAGCTGA 549
 859 pGluThrLysLeuAsnThrValAspAspGlnLysAlaGlySerProSera 876
 550 CGAGACTAAACTCAATCAGTGTGATGCCAGAAAGCGTGTCTCCAGCA 599
 876 rGAspValGlyProSerLeuGlyLeuLysLysSerSerSerLeuGluSer 892
 600 GAGATGTGGTCCCTCCCTGGGTCTGAGAGAGTCAGCTGTTGGAGAGT 649
 893 LeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIleProPheH1 909
 650 CTGACAGCCGACAGTGTGCGAGGTGACTTTGATGGGATATCTCTTCCA 699
 909 sArgProArgProArgGlyIleArgGlyArgGlyCysAsnGluSerPhea 926
 700 TCCTCCACGGCCCGGATTAATCAGAGGCGAGGTGATGCAATGAGACTTCA 749
 926 rGAlaAlaAlaIleAspLysSerTyrAspLysProAlaValAspAspAsp 942
 750 GAGCTGCATCGACAAATCTTATGATTAACCCGGGTAGATGATGATGAT 799
 943 GluGlyMetGluThrLeuGluGluAspThrGluGluSerSerArgSerG1 959
 800 GAAGGCGATGAGACCTTGGAAAGACACAGAAAGAAAGTTCAAGATCAGG 849
 959 YArgGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGluA 976
 850 GAGAGAGTCTGTATCCACAGCCAGTGTACGCTTCCACTCTCTTGAGA 899
 976 rGgInMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArgLysLys 992
 900 GACAAATGAATCGAAACCAAGAGAAAGGTATTAAGACTGATAGAAAAG 949
 993 AspLysThrGlyLysGluLysLysLysAspArgAspLysGluLysAspLys 1009
 950 GATAAAACTGAAAAGAAAAGAAAGATAGAGATAGAGAGAAAGGATAA 999
 1009 sMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgP 1026
 1000 AATGAAAGCCAAAGAGGAATGCTGAAAGGCTTGGGAGACATGTTCAGGT 1049
 1026 heGlyLysHisArgLysAspAspLysIleGluLysThrGlyLysIleLys 1042
 1050 TTGGCAAACTCCAAAAGATGACAGATTGAGAAAACGGTAAATATAAA 1099
 1043 IleGlnGluSerPheThrSerGluGluGluArgIleArgMetLysGlnG1 1059
 1100 ATACAGGAATCCTTTACATCAGAAAGAGAGGATACGAATGAAGCAGGA 1149
 1059 uGInGluArgIleGlnAlaLysThrArgGluPheArgGluArgGlnAlaA 1076

1150 GCAGAGAGGAGATTCACGCAAAACTCGAGAAATTATGGGAGACAGCAAGCTC 1199
 1076 rGgLuArgAspTyrAlaGluIleGlnAspPheHisArgThrPheGlyLys 1092
 1200 GAGAGCGTGACTATGTGAATTCAGATTTTCAGATCGGACATTTGGCTGT 1249
 1093 AspAspGluLeuMetTyrGlyGlyValSerSerTyrGluGlySerMetAl 1109
 1250 GATGATGATTAATGATATGGGGAGTTCTCTTATGAAAGTTCCATATGC 1299
 1109 aLeuAsnAlaArgProGlnSerProArgGluGlyHisMetLeuAlaL 1126
 1300 TCTCAACCTAGACCTCAGAGCCACGAGAGGGCATGTGATGATGCTT 1349
 1126 eUtyrSerPheGlnValLysLysProArgAsnSerLysProSerProValAsp 1142
 1350 TGATATGCCCAAGTCAGAGACCGCGGAATTCCAAACTCCACCTGTATGAC 1399
 1143 SerAsnArgSerThrProSerAsnHisAspArgIleGlnArgLeuArgG1 1159
 1400 AGTAACAGATCAACTCTAGCAATCATGATCGATACAGCGCTGTGAGCA 1449
 1159 nGluPheGlnAlaLysGlnAspGluAspValGluAspArgArgArgT 1176
 1450 AGAATTTTCAGCAAGCAAGCAAGATGAAGATGTAGATGATGCTGCCGA 1499
 1176 hrTyrSerPheGlnGluGlnProTrpProAsnAlaArgProAlaThrGlnSer 1192
 1500 CTTATATGTTTGAGACMACCTTGCCGAGAGCGACGGCCGACGAGAGC 1549
 1193 GlyArgHisSerValSerValGluValGlnMetGlnArgGlnArgGlnG1 1209
 1550 GGGGAGACACTGGGTGTCCTGGAGGTGAGATGACAGCGGCGGACGAGCA 1599
 1209 uGluArgGluSerSerGlnGlnAlaGlnArgGlnTyrSerSerLeuProA 1226
 1600 GGACGCGCAGAGCTCCACGAGCGCCACGCGCAGTACAGCTCTGCGCTC 1649
 1226 rGgInSerArgLysAsnAlaSerSerValSerGlnAspSerTrpGluGln 1242
 1650 GGCAAGAGCAGAAAATATCCAGCTCGTCTCCAGAGACTTGGAGCAG 1699
 1243 AsnTyrSerProGlyGluGlyPheGlnSerAlaLysGluAsnProArgTyr 1259
 1700 AACTACTCCCTGGGGAAGGCTTCCAGAGTGCCTCAAGAGAAACCCAGGTA 1749
 1259 rSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyGlyHisGlyPhea 1276
 1750 CTCAGAGTACCAAGGCTCCAGGAAGCGCTACCTGGAGAGACATGGCTTCA 1799
 1276 snAlaArgValMetLeuGlnThrGlnGluLeuLeuArgGlnGluGlnArg 1292
 1800 ACGCCAGGGTATGCTGGAACCTCAGAGCTCTTCCAGAGAAACAGAGG 1849
 1293 ArgLysGluGlnGlnMetLysLysGlnProProSerGluGlyProSerAs 1309
 1850 CGGAAGGAGCAGCAGATCAAGAAAGCAGCTCTCCGAGGGGCCCCAGCA 1899
 1309 nTyrAspSerTyrLysLysValGlnAspProSerTyrAlaProPolysG 1326
 1900 CTATGACTCGTATAGAAAGTCCAGGAGCCCAAGTACCCCTCCCAAGG 1949
 1326 LyrProPheArgGlnAspValProProSerProSerGlnValAlaArgLeu 1342
 1950 GGCCCTTCCGCAAGATGTGCCCCCTCTCTCTTCAGAGTTGCCAGGCTG 1999
 1343 AsnArgLeuGlnThrProGluLysGlyArgProPheTyrSer 1356
 2000 AACAGACTTCAGACTCTGAGAAAGGAGGCCCTTCTATTTCC 2041

seq_name: gb_pr.AF332592

seq documentation block:

LOCUS AF332592 2967 bp mRNA linear PRI 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein
short variant b mRNA, complete cds, alternatively spliced.
ACCESSION AF332592
VERSION AF332592.1 GI:14579308
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2967)
AUTHORS Fang, C.M. and Xu, Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
asip isoforms in human hepatocellular carcinomas
JOURNAL Cell Res. 11 (3), 223-229 (2001)
MEDLINE PUBMED 11642408
REFERENCE 2 (bases 1 to 2967)
AUTHORS Fang, C. and Xu, Y.
TITLE Exon/Intron Structure And Splicing Variants of A Novel Human
Polarity Gene, hasip
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2967)
AUTHORS Fang, C. and Xu, Y.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
Oncology, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
320 Yue-Yang Road, Shanghai 200031, P. R. China
FEATURES
source
1..2967
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..2967
/note="ASIPSB; alternatively spliced; similar to GenBank
Accession Number AF196186"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein short variant b"
/protein_id="AAK69192.1"
/db_xref="GI:14579309"
/translation="MKVTVCFGRTNVPVCGGHNKVFSLIQAVTRKRIADPNY
WIVNRLHEGGGILDLDDILCDVADKDRIVAVDEQDPHGGDGTASSTGOSPE
IFSEELGNNVSAFOPYOATSEIETPSVLNAMPILHYRRSDPALIGLSTYSDSNF
SSEPSRKNPTFRSTTACFTLAKONTAGSPKTCDRKDEYRGLPHDTNMSNOQRDA
RSLSSHPMCKWLEKOBDEGTEEDNSREVEPVGKDEHLEHITPMSLDDWKLVE
VPMDDGPLCIVHVPFSARGRTLGLVRLKEGKAHEENLPENDCTVIRINDGLIN
RRERQKHFRQAMRTPIWHVPAANKREQEOLSGSEKNNTYSSRFSDQYIDNR
SYNSGLHTVQAPRLNHPQIDSHSRILPHSARSPKPSAPASAPQNVSTTVSSG
ITWKTKGRKLNIOLEKKTGEGFGFSTTSRDYTGSAATLVKNLIPKGAALODRLKAG
DRLIENGCVDLVGKSOEVEVSLRSTKEGTVALVPRQDAPHPHELAEADIDVLT
PDGREPLTFEVLNDSSAGLGVSKNSKSKENHMLGLFVNSIINGCAAGDGGTR
VNDQLIANGESLIGKTNOMAMETLRKSMSTENKGMQLIIVARISMCNLKSPGS
PPGELPIETALDPERIRHSISYSGIEGLFVADGTAKAKIASAPACSTVPPNRPQD
DTVITIEDRFPVLPPLHSDSSSHDDVGFVADGTAKAKIASAPACSTVPPNRPQD
VIATOREFGQIADETKLTNTVDOKGSPRSGLTKSSSLESQTLVAAPVTL
NGDIPFRHPRRIIRGRGNEFRRAIDKSYDKPAVDDDEMETULEETDESSSGR
ESVYASDQSHSLDEROMNGNOKGDKTDKKKDKTKEKKKDRDEKDKMAKKGMKLK
GLDMFSLAKIKPEKR"

BASE COUNT 890 a 653 c 776 g 648 t
ORIGIN

alignment_scores:

Quality: 556.00 Length: 556
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:

US-09-757-781-2 x AF332592 ..

Align seq 1/1 to: AF332592 from: 1 to: 2967

```
1 MetLysValAlThValCysPheGlyArgThrArgValAlValProCysG1 17
|||||
1 ATGAAGATGACCGTGTCTTCGAGAGGACCCGGTGGTCCGCGCG 50
17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
|||||
51 GGACGGCCCATGAAAGATTTCACGCTCATCCACAGCGGTGACCCGCT 100
34 yArgGlyAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
101 ACCGGAGGCGCATGCCAGATCCAAACTACTGATACAGGTGCATCCG 150
51 LeuGluHisGlyAspGlyIleLeuAspLeuAspAspIleLeuCysAs 67
|||||
151 TTGGACATGAGATGAGAGCAATCTAGACCTTGATGACATTTGTGGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
|||||
201 TGTAGCAGACGATAAAGACAGACTGTGACAGTGTATGACAGAGATC 250
84 rHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
251 CACATCAGGAGGTGATGGCACCATGTCACAGTCCACGCGTACCCAGAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
|||||
301 CCAGACATATTTGTGTAGTACCTGGCCACCAACATCTCCACGCTTTCA 350
117 nProGlyGlnAlaThrSerGluIleGluValAlaThrProSerValLeuArg 134
|||||
351 GCCTTACCAACCAACAGTGAATTTGAGTCACAGCTTCAGTCTTCGAG 400
134 IaAsMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
|||||
401 CAAATATGCCCTTCATGTTGACGAGCAGTAGTACCCACAGCTCTAATTGCC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGlnProSerArg 167
|||||
451 CTCCTCCACTTCTGATGATAGTATTTTCTCTGAAAGACCTTCACAG 500
167 gLysAspProThrArgTrpSerThrAlaGlyPheLeuLysGlnAsnT 184
|||||
501 GAAAAATCCACACGCTGTTCAACAACAGCTGGCTTCTCAAGAGAAACA 550
184 hTrIaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
|||||
551 CTCTGGAGTCTTAAACCTGCGACAGAAAGAAAGTAAACTACAGA 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
|||||
601 ACCCTCCCGCGGATACTAGTACCTGCTTAAACCAATTTACAGAGACAA 650
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
|||||
651 TGTCTGCTGCTCTGTGATGCGAGTCAACCAATGTTGGCAGATGCTGG 700
234 LuLysGlnGluGlnAspGluAspGlyThrGluGlnLysAsnSerArgVal 250
|||||
701 AGAAACAAGAAACAGATGAGATGCGACAGAAAGAGATACAGTGTGTT 750
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheSe 267
|||||
751 GAACCTTTGACATGCTGACACGGGTTTGGACATATACCAACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
|||||
801 TCTGATGATATGTTAAAGCTCTGTAAGATCCCAACGATGAGGCGCTC 850
284 euGlyIleHisValAlaProPheSerAlaArgGlyArgThrLeuGly 300
|||||
```

```

851 TGGGAATCCATGATGATGCTTTCAGTCTCGAGCGCAGACCCCTGGG 900
301 LeuLeuValIysArgLeuGluIysGlyIysAlaIuHisGluAspLe 317
|||||
901 TTATTAGTAAAGCATTTGGAGAAAGGTGTAAAGCTGAACTGAAACT 950
|||||
317 uPheArgGluAsnAspCysIleValArgIleAsnAspCysIleuArg 334
|||||
951 TTTTCGTGAGAAATGATGATGATGATGATGATGATGATGATGATG 1000
|||||
334 snArgArgPheGluGluAlaGlnHisMetPheArgGlnAlaMetArg 350
|||||
1001 ATTAAGATTTGTAACACACACATATATGTTTCCGCAAGCATCGTAC 1050
|||||
351 ProIleIleTrrPheHisValIleProAlaIleAsnIysGluGlnTyr 367
|||||
1051 CCCATCATTTGTTCCATGTTGTTCTCCACCAATTAAGAGCACTATGA 1100
|||||
367 uGlnLeuSerGlnSerGluIysAsnAsnTyrTyrSerSerArgPheSer 384
|||||
1101 ACAACTATCCCAAGTAGAGAAACAATTAATTAATTAATTAATTAAT 1150
|||||
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
|||||
1151 CTGACAGCCAGTATATGACACAGAGAGTGTGAACAGTGCAGGCTTAC 1200
|||||
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAsp 417
|||||
1201 AGGTGCGAGAGAGACCCCGACTGACACCCGCTGAGAGATAGACTC 1250
|||||
417 HisSerArgLeuProHisSerAlaHisProSerGlyIysProProSer 434
|||||
1251 TCACCTCAGACTACCTCATGACGACACCCCTCGGGAACACCATTCG 1300
|||||
434 IapProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
|||||
1301 CTCAGGCTCGGACCTCAGATGATATTAATGACACTGTAAAGAGTGT 1350
|||||
451 TyrAsnThrIysIysIleGlyIysArgLeuAsnIleGlnLeuIysGly 467
|||||
1351 TATTAACACCAAAATATGCAAGAGGCTTAATATCCAGCTTAAGAAAG 1400
|||||
467 YThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
|||||
1401 TACGAGAGGTTGGGATTCAGATCATTCCAGAGATTAACAATAGGTG 1450-
|||||
484 IysSerAlaProIleTyrValIysAsnIleLeuProArgIleAlaIle 500
|||||
1451 GCTCAGCTCCAACTATATGTAATAACATTTCCCGGGGGGGGCGCAT 1500
|||||
501 GlnAspGlyArgLeuIysAlaGlyAspArgLeuIleGlnValAsnGly 517
|||||
1501 CAGATGCGCAGCTTAAGAGAGAGAGACTTATAGAGTAAATAGAGT 1550
|||||
517 IAspLeuValIleIysSerGlnGluValIleValSerLeuLeuArgSer 534
|||||
1551 AGATTATGCGCAATCCCAAGAGAGAGTGTTCGCTGTGAGAACCA 1600
|||||
534 hTrIysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAsp 550
|||||
1601 CCAAGATGAGAAAGAGAGTGTGAGCTTGTGTCGTCGAGAGAGAGCC 1650
|||||
551 PheHisProArgGluLeu 556
|||||
1651 TTCCAGCCCAAGGAACTG 1668
|||||
seq_name: gb_pr:AF332593
seq_documentation_block:
LOCUS AF332593 3822 bp mRNA linear PRI 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isoform-specific interacting protein long
variant b mRNA, complete cds, alternatively spliced.
ACCESSION AF332593

```

```

VERSION AF332593.1 GI:14579310
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 3822)
Fang, C. M. and Xu, Y. H.
Down-regulated expression of atypical PKC-binding domain deleted
asp isoforms in human hepatocellular carcinomas
JOURNAL Cell Res. 11 (3), 223-225 (2001)
MEDLINE 21520080
PUBMED 11642408
REFERENCE
AUTHORS 2 (bases 1 to 3822)
Fang, C. and Xu, Y.
Exon/Intron Structure And Splicing Variants of A Novel Human
TITLE Polarity Gene, hASIP
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 3822)
Fang, C. and Xu, Y.
Direct Submission
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
Oncology, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
320 Yue-Yang Road, Shanghai 200031, P. R. China
FEATURES
source
1..3822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..3822
/note="ASIPB: alternatively spliced; similar to Genbank
Accession Number AF196185"
/codon_start=1
/product="atypical PKC isoform-specific interacting
protein long variant b"
/protein_id="AAK69193.1"
/db_xref="GI:14579311"
/translacion="MKVYCGFGRVYVPCGDDHMKVFSLIQAVTRRYAKDPY
WIVHLEHEDGGILIDDLICDVAADKDLVAFEDPDHGGDGTASCTGSPF
IEGSELTNNVSAFOPVQATSELEVPSVLRAKMPHVRSSPALIGLSTVSDSF
SSEPSRKNPTKMTSTAGFLKQNTAGSPKTCDDKKEENYSLPRDYSNMQVQDRA
RSSLSASHPVQKMLEKODEDEGTEEDNSRVEPVGADGLHIPFSLDDWKLVE
VPNDGPIGLIHVVPFSARGRTGLLVKRLKGAKEHGLFRENDCIVILINDGLNR
RREFOAHMEFROAMPPIIWEHYVPAANKROYOLSOSEKNYSSRFSPDSQIDNR
SVNSAGLHTVORAPRLNHPPEOIDSRLPHSAHSGKPSASAPASAPONFSTVSSG
YNTKIKGRLNIDLKGTEGLGFSITSRYDTIGSAPITYKNILPRCAIIDGRLAG
DRLIEVNGVDLVKRSOEYVSLRSTKMEGTVELVFRQDAPRPLKEDDVIAT
PIDTRFELFEVPLNDGSAAGLVSVKNSKKNHDLGIFVSIINGAASDGRIR
VNDQLAVNGESILGKTNDAMETLRSKSTENKRMGLIYARISKNEKSPGS
PPGPELPIETALDDERRISLSYSGIEGDESPRSALSRIMGKTLQSLPTNMPDQ
DVIITDDRLPIVPHLSDOSSSHDDVGFVAADGTAKKAISDADCSLSPDNP
VIAFORGEGROJADETKLNTVDDKAGSSRDVPSGLKSSLSLESLDTAYAEVTL
NGDIPRHPRPRIILRGKCNESFRALIDKSYDPAVDDDEGTELEDEESSRSGR
ESVSTASIDOPSHLEKQNGNOKGKTDKDKTKCEKKDKDKEDKMKAKGMLK
GLDLMDFRIQAKTREFEERQARERDYAIEIOPHRTFGCDDIELAGVSSYSGALNR
POSPREGHMADALYAYQKRPNSKSPVDSNRSTPSNHDRIOIRLROFOOAKDVE
DRRRTYSFEQPMNPARNAPATQSGHSVSEVOMOROREEESQOAGROVSSI.PROS
KNASYSQDSWENQNTSPGEGFQAKENPRISSTIGSNGTGLGCHGNARMLTQELL
ROBORREDOQMKRQPSSEGPSNTDSTIKVODPSIAPKGFPRDVPSPQAVAKNLRL
QTPKGRFPYS"
BASE COUNT 1114 a 885 c 1019 g 804 t
ORIGIN
alignment_scores:
Quality: 556.00 Length: 556
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AF332593

```

Align seg 1/1 to: AF332593 from: 1 to: 3822

```

1 MetLysValThrValCysPheGlyValArgValValValProCysG1 17
1 ATGAAGATGACCGTCTCTCTGCGACGGACCGGGGCGTGGCCGCGG 50
17 AspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArg 34
51 GGACGGCCACATGAAGATTTCAGCTCATCCACAGCGGTGACCCGCT 100
34 YrArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
101 ACCGGAAGGCCATCGCCCAAGATCCAAACTAGTATACAGATGACATCG 150
51 LeuGluHisGlyAspGlyGlyIleLeuAspLeuAspPheIleLeuGly 67
151 TTGGAAATGAGATGAGAGAAATCTAGACCTTGATGACATTTCTTTGTA 200
67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGluAspP 84
201 TGTAGCAGACGATTAAGACACAGACTGGTAGCAATGTTGATGACAGATC 250
84 TrHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
251 CACATCACGGAGGTGATGGCACCAAGTCCAGTCCAGGGGTACCCAGAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
301 CCAGACATATTTCGTAGTAGCTGGCACCAAAATGTTCTCAGCCTTCA 350
117 nProGlyGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAAGCAACAGAAATTGAGGTACACCTTCAGCTTCGAG 400
134 LaAsnMetProLeuHisValArgSerSerAspProAlaLeuIleGly 150
401 CAAATATGCTCTTCATGTTCACCCAGATGAGCCAGCTCTAATGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSer 167
451 CTTCTCCACTTCGTCTGATGATGATTAATTTTCCCTGAGAGCCTTCA 500
167 GlysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsn 184
501 GAAAAATCCACACGCTGTGTCAACACAGCTGCTCTCAAGCAGACA 550
184 hAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTGCTGGAGAGTCTAAACCTGCGACAGAAAGTAAGTAACCTACAGA 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
601 AGCTCCCGCGGGATTAAGTAACTAGTAACTAGTAACTTCCAGAGACA 650
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeu 234
651 TGTCTGCTGCTCTGTAGTCCAGTCAACCAATGAGGCAAGTGGCTG 700
234 LuysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
701 AGAAACAGAAACAGATGAGATGGAGCAGAAAGAGATTAACAGTGTGT 750
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhe 267
751 GAACCTGTTGACATGCTCACACGGCTTTGGAGCATATACCAACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysPro 284
801 TCTGGATGATATGTAAGCTCTGTAAGTCCCAAGATGGAGGCTTC 850
284 euGlyIleHisValValProPheSerAlaArgGlyGlyArgThrLeuGly 300
|||||

```

```

851 TGGGAATCCATGATGAGTCCCTTCACTGCTCGAGCGCGAGAACCTCGGG 900
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLe 317
901 TTATTATGTAAGAAAGATTGAGAAAGTGGTAAAGTGAACATGAATACT 950
317 uPheArgLysAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
951 TTTTGTGATGAATGATTTTCATTTGTCAGATTAATGATGGGACCTCGAA 1000
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1001 ATAGAAATTTGAACACAGACACATATGTTTCCCAAGCCATGCTTACA 1050
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluIntArg 367
1051 CCCATCATTTTGTGTTCCATGTTGTTCTGTCAGCAAAATTAAGCAGTATGA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTrpSerSerArgPheSerP 384
1101 ACAACTATCCCAAGTGAAGAAACAATTACTATTCAGCCGTTTAGCC 1150
384 rAspSerGluTrpIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1151 CTGACAGCCGCTATATTGACAAACAGAGAGTGTGAACAGTCCAGGCTTAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAsp 417
1201 ACGGTCAGAGACACCCGACATGAAACACCCGCTGACAGCATAGACTTC 1250
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSer 434
1251 TCACCTAAACACTACCTCATATGACACACACCCCTGGGAAACACCATCG 1300
434 lAProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1301 CTCACGCTCCGACACTCAGATGATTAATTAACACTGTAAACAGATGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGly 467
1351 TATTAACCCMAAAATATGCAAGAGGCTTAATATCCAGCTTAGAAGG 1400
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
1401 TACAGAAAGTTGGGATTCACATCATCTCCAGAGATGTAAACATAGTGT 1450
484 lYserAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1451 GCTACGCTCCATCTATGTGAAACATTCCTCCCGGGGCGGCCCATTT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGly 517
1501 CAGGATGGCGACTTAAGCAGAGACAGACTTAATAGAGTAATGAGAT 1550
517 lAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534
1551 AGATTATGAGGCAAAATCCCAAGAGAAAGTTGTTGCTGTTGAGAGACA 1600
534 hTrpMetGluGlyThrValSerLeuValPheArgGlnGluAspAla 550
1601 CCAAGATGGAAGAACTGTAGAGCTTCTGTCTTTCGCCAGGAAGAGCC 1650
551 PheHisProArgGluLeu 566
1651 TTCACCCCAAGGAACTG 1668

```

seq.name: qb_pr:AF467006

seq.documentation_block:

LOCUS AF467006

DEFINITION Homo sapiens partitioning-defective 3 protein splice variant f (PARD3) mRNA, complete cds, alternatively spliced.

ACCESSION AF467006

3933 bp

mRNA

linear

PRI 07-FEB-2002

```

VERSION      AF467006.1  GI:18568351
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 3933)
AUTHORS      Gao, L., Macara, I.G. and Joberty, G.
TITLE        Multiple splice variants of PAR3 and of a novel related gene,
              PAR3l, produce functionally different proteins
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 3933)
AUTHORS      Gao, L., Macara, I.G. and Joberty, G.
TITLE        Direct Submission
JOURNAL      Submitted (10-JAN-2002) Center for Cell Signaling, University of
              Virginia, PO Box 800577, Charlottesville, VA 22908, USA
FEATURES
source       1..3933
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
gene         1..3933
              /gene="PAR3"
              /note="PAR3"
CDS          1..3933
              /gene="PAR3"
              /note="PAR3f: multi-PDZ domain containing protein;
              alternatively spliced"
              /codon_start=1
              /product="partitioning-defective 3 protein splice variant
              f"
              /protein_id="AAL76046.1"
              /db_xref="GI:18568352"
              /translation="MKVTCFGRTRVVPVPCDGHMKVFSLIQQAVTRYRAIKADBNY
              WIOVHLEHGDGIDLDLDDILCDVADDDKDLVAVPEODPHHGDDGTSASTGSGE
              TSGSELTNNVSAFOPYOATSEIEVPSVLRAAMPILVRRSSDPAIGISTSDSDNE
              SSEPSRKRPSTKSTAGFLKONTASPTCKRKENTRSLEPRDSNMSNOQRDPA
              RSLASHPVAGKMLEKQODEDETEDSRVEPVHADGLSHINFSLDNKLVE
              VPNDGGPILIHVVPFSAKGRITGLLYKLEKGAKEHLENFENDCIYVINGDLN
              RRFEOAHMFROMRTPITMFWVPANKEOYEQOLSOSENNYSSRFPDSOYIDNR
              SVNSAGLHTVORAPRLNHPPEQIDSHSLPHSAHPGKPPASAPASAPONVESTYDNR
              VYMKIKGRRLIOLKKGTEGLGFSITSRDVTIGGSAPIYVKNILPPGAALODRLKAG
              DRLIENGVDIVKSGOEYVSLIRTKMGTVSLIYFROEDAHPRELAEDDLYLT
              PGTRFELTFEVPPLNDGSGAGLVSYGNRSKSNADLGIYVASTIINGNAGSGLR
              VNDQLAVGESLGLKTNODAMETLRSMSTEBKNGMQLIVARRISKNELKSGS
              PGPELPILFTALDDRRRISHSLYSGIEGLDESPNNALISRIKRYQLSPYNNMPOD
              DVIILEDRLPYLPHLSDSSSSSHDDVGFYADAGTAKAKAISDAPCSLSDVDP
              VLAFOREGFGROIADETKLNTVDOKAGSPSRVPSLIGIKSSLSLEOTAVAEVTL
              NGDIPHRPRPRLIRGRCNESFRALIDKSYDKPAYDDDDDEGNETLEEDTEESSRGR
              ESYVTSADSPSHSLEROMKNGNOKGDKTDKKKTEKKKDKDKKAKKMKL
              GLGDMEFRGKHRRDKLERTGKIKIQESFSEBERLRMOQEDRIAKTREFEQRAR
              ERDYAEIODEHRTFGCDDELMYGGVSYEGSMALNARPOSFRGHMMADALYAQVKPR
              NKSPPVDENRSTPSNHDRIQLRLOFEQAKODEVDYRRRTYSFQPMNAPAROS
              GRHSVVEVOMOROREERESSQAOAROYSSLPROSRKNASVSDOSWDRNYPDGGF
              QSAKEPRYSSTYOGSRNGYLGIGFNARMALETOELROEQRKEOMKQKPSSEGS
              NVDSTIKVDDPSTAPKPPFRODVPSPSOVAVNLRLQTEPRKPRYS"
              814..1074
              /gene="PAR3"
              /note="Region: PDZ domain"
              1378..1638
              /gene="PAR3"
              /note="Region: PDZ domain"
              1726..1989
              /gene="PAR3"
              /note="Region: PDZ domain"
              /note="Region: PDZ domain"
BASE COUNT  1164 a      899 c      1049 g      821 t
ORIGIN
alignment_scores:
  Quality:      556.00      Length:      556
  Ratio:        1.000      Gaps:        0
  Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-09-757-781-2 x AF467006
Align seg 1/1 to: AF467006 from: 1 to: 3933
1 MetLysValThrValCysPheGlyArgThrArgValAlaValProCysGI 17
1 ATGAAAGTGACCGTGTGCTTCGACGAGCAACCCGGGTGCTGCTCCGTGGCG 50
17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
51 GGACGGCCACATGANAAGTTTTCAGCCTCATCCAGAGGGGGTGACCCGCT 100
34 yrArgLysAlaIleAlaLysAspProAsnYrTrpIleGlnValHisArg 50
101 ACCGGAAGCCATCCGCAAGATCCAAACTACTGATACAGGTGCATGCATGC 150
51 LeuGluHisGlyAspGlyGlyIleLeuAspLeuAspAspIleLeuCysAs 67
151 TTGGAAACATGGAGATGAGGAATACTAGACCTTGATGACATTCCTTTGTGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAspP 84
201 TGTACAGACGATAAAGACAGACTGATGACAGTGTGTTGATGACAGAGATC 250
84 rHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
251 CACATCAGAGGAGTGATGACACACAGTGCACAGTCCACAGGATCCAGAC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheGI 117
301 CCAGAGATATTTGTTGATGAGTGTGACCAACAATGTTCTCAGCCCTTCA 350
117 nProGlyGlnAlaThrSerGluIleGluValThrProSerValIleuAsp 134
351 GCCTTACCAAGACACAGGAATGAGGTCAACCTTCATGCTTCGCGAG 400
134 IaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
401 CAATATGCGCTTCATGTTGACAGCAGTAGACACCACTTANTTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
451 CTCCTCACCTTCGTGATGATAGTATTTCTCTGTAAGACCTTCAG 500
167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAspT 184
501 GAAATATCCACACGCTGCTCAACACACAGCTGCTTCTCAAGACGAACA 550
184 hrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnYrArg 200
551 CCGCGGGAGCTCTAAACCTCGACAGAGAAAGATGAAAACATACACA 600
201 SerLeuProAlaGAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
601 AGCCCTCCGCGGGATACTAGTAACTAGTGTAAACCAATTTCACAGAGACA 650
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
651 TGTCTGCTGCTCTGTGAGTGCAGTCACCAATGGTGGGCAAGTGGCTGG 700
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluLysAspAsnSerArgVal 250
701 AGAAACAGAAACAGATGAGATGGGAGACAGAAAGATATACAGTGTGTT 750
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheSe 267
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284
801 TCTGATGATATGTTAAGCTCTGTAGAAAGTCCCAACGATGAGAGGGCTTC 850

```

seq_name: qb_pr:AF252293	
seq_documentation_block:	
LOCUS AF252293	3801 bp
DEFINITION Homo sapiens partitioning-defective 3 splice variant c (PAR3) mRNA.	linear
	PRI 10-JAN-2002

```
misc_feature 682. .942 /gene="PAR3" /note="Region: PDZ domain 1"
misc_feature 1246. 1506 /gene="PAR3" /note="Region: PDZ domain 2"
misc_feature 1594. 1851 /gene="PAR3" /note="Region: PDZ domain 3"
```

```

BASE COUNT      1121 a      867 c      1016 g      797 t
ORIGIN

alignment_scores:
    Quality: 500.00      Length: 500
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x AF252293      ..

Align seg 1/1 to: AF252293 from: 1 to: 3801

857  ILAAlaSpGLuThrLysLeuAsnThrValAspAspGlnLysAlaGlyse 873
|||||
2299 ATAGCTGACGAGCTAAACTCAATACAGTGTGATGACCGAAAGAGGTTTC 2348
|||||
873  rProSerArgAspValGlyProSerLeuGlyLeuLysLysSerSerSert 890
|||||
2349 TCCGAGCAGAGATGTGGTCTTCTCCGTGGGTCTGAAAGAGCTCGT 2398
|||||
890  euGluSerLeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIle 906
|||||
2399 TGGAGAGCTTCGACAGCCGACGCTCCGAGGTGACTTTGAAATGGGGAATAT 2448
|||||
907  ProPheHisArgProArgProArgIleIleArgGlyValArgGlyCysAsnG 923
|||||
2449 CCTTCCATGCTCCACGGCCGCGGATTAATCAGAGCCAGGGGATCCAATGA 2498
|||||
923  uSerPheArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspA 940
|||||
2499 GAGCTTCAGAGCTGCCATCGACAAATCTTATGATAAACCCGCGGTAGATG 2548
|||||
940  spAspAspGluGlyMetGluThrLeuGlnLysAspThrGluGluSer 956
|||||
2549 ATGATGTAGTGGGCGATGAGACCTTGGAGAGACACAGAAAGAAAGTTCA 2598
|||||
957  ArgSerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSe 973
|||||
2599 AGATCAGCGAGAGAGCTGTATCCACAGCCAGTATCAGCCTTCCACTC 2648
|||||
973  rLeuGluArgGlnMetAsnGlnLysGlnLysGlyAspLysThrAspA 990
|||||
2649 TCTGGAGAGCAAAATGATGAAACCAAGAGAAAGTGATTAAGACTATA 2698
|||||
990  rGlySLysAspLysThrGlyLysGluLysLysLysAspArgAspLysGlu 1006
|||||
2699 GAAAAAAGGATAAACTGGAAAAAGAAAAAGAAAGATAGAGATPAAGAG 2748
|||||
1007  LysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMe 1023
|||||
2749 AAGGATTAATTAAGAAAGCCAGAAAGGAAATGCTGAAGGCTTGGAGACAT 2798
|||||
1023  tPheArgPheGlyLysHisArgLysAspAspLysIleGluLysThrGlyL 1040
|||||
2799 GTTCAGGTTGGCAAAACATCGAAAAAGATGACAAACATTGAGAAAAACGG 2848
|||||
1040  ysIleLysIleGlnGluSerPheThrSerGlnGluGluValArgMet 1056
|||||
2849 AATATAAAATATACAGGAATCCCTTACATCAGAGAGAGAGAGATACGAATG 2898
|||||
1057  LysGlnGlnGlnGluArgIleGlnAlaLysThrArgLysPheArgGluAr 1073
|||||
2899 AAGCAGGAGCGAGAGAGATTCAGACCAAAACCTGAGAAATTAAGGAACG 2948
|||||
1073  gGlnAlaArgGluArgAspTyrAlaGluIleGlnAspPheHisArgThrP 1090
|||||
2949 ACAAGCTCGAGAGCGTCACTATCTGAAATCAAGATTTTCATCGGACAT 2998
|||||
1090  heGlyAspAspAspGluLeuMetTyrGlyGlyValSerSerTyrGluGly 1106
|||||
2999 TTGGCTGTGATGATGAGTTAATGTATATGGGAGGATTTCTTCTTATGAAG 3048
|||||

```

```

1107  SerMetAlaLeuAsnAlaArgProGlnSerProArgGlnGlyHisMetMe 1123
|||||
3049 TCCATGGCTCTCAACGCTAGACCTCAGAGCCACAGAGAGGCGATATGAT 3098
|||||
1123  tAspAlaLeuTyrAlaGlnValLysLysProArgAsnSerLysProSerP 1140
|||||
3099 GGATGCTTTGTATGCCCAAGTCAGAAAGCCGCGCAATTCCAAAACCTTCAC 3148
|||||
1140  roValAspSerAsnArgSerThrProSerAsnHisAspArgIleGlnArg 1156
|||||
3149 CTGTAGACAGTAAACAGATCAACTCTTACCAATCATGATGGATACAGCT 3198
|||||
1157  LeuArgGlnGlnPheGlnGlnAlaLysGlnAspGlnLysAspValGluAsp 1173
|||||
3199 CTGAGCGCAAGAAATTTACAGCAAGCAAAACAGATGAAGATGAGAAAGTCG 3248
|||||
1173  gArgArgThrTyrSerPheGlnGlnProTyrProAsnAlaArgProAlat 1190
|||||
3249 TCGCGGACCTATAGTTTTCAGCAACCTCGGCGCAGACGCGCGCGA 3298
|||||
1190  hrgInserGlyArgHisSerValSerValGluValGlnMetGlnArgGln 1206
|||||
3299 CGCAGACCGGGCGACACTCGGTGTCGTGGAGGTGCAGATCACCGGCGAG 3348
|||||
1207  ArgGlnGlnGluArgGluSerSerGlnGlnAlaGlnArgGlnTyrSerSe 1223
|||||
3349 CGGAGAGAGAGAGCGCGAGAGCTCCACAGAGCCACAGCGCCAGTACAGCTC 3398
|||||
1223  rLeuProArgGlnSerArgLysAsnAlaSerSerValSerGlnAspSerT 1240
|||||
3399 TCTCTCTCGCGCAAGACAGGAAATAATGCCAGCTCCGTCTCCAGACACTCT 3448
|||||
1240  rPglGlnAsnTyrSerProGlyGlnGlyPheGlnSerAlaLysGlnAsn 1256
|||||
3449 GGGAGCAGAACTACTCCCTGGGAAGGCTTCCAGAGTCCAAAGAGAGAAC 3498
|||||
1257  ProArgTyrSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyGlyH 1273
|||||
3499 CCCAGGTACTCTCAGACTACCAAGGCTCCAGGAAGGCTACCTGCGAGGACA 3548
|||||
1273  sGlyPheAsnAlaArgValMetLeuGluThrGlnGluLeuLeuArgGln 1290
|||||
3549 TGGCTTCAACCCCAAGGCTCATGTGAAACTCAGAGAGCTCTTCCGAGG 3598
|||||
1290  LglInArgArgLysGlnGlnMetLysLysGlnProProSerGlnGly 1306
|||||
3599 AACAGAGCGGAGAGAGCGAGATGAAGAAAGCAAGCTCTTCGAGGGG 3648
|||||
1307  ProSerAsnTyrAspSerTyrLysLysValGlnAspProSerTyrAlaPr 1323
|||||
3649 CCCAGCAACTATGACTGCTATTAAGAAAGTCCAGAGACCCCAATTACGCCCC 3698
|||||
1323  oProLysGlyProPheArgGlnAspValProProSerProSerGlnValA 1340
|||||
3699 TCCCAAGGGCGCTTCCGCGCAAGATGTCCCCCTCCCTCTCATGAGGTTG 3748
|||||
1340  laArgLeuAsnArgLeuGlnThrProGluLysGlyArgProPheTyrSer 1356
|||||
3749 CGAGGCTGAACAGACTTCAGACTCTCGAAGAAAGGAGGCCCTTCTATATCC 3798
|||||

seq_name: gb_BCO11711

seq_documentation_block:
LOCUS      BCO11711                      1592 bp      mRNA      linear      PRI 02-AUG-2001
DEFINITION Homo sapiens, clone MGC:19518 IMAGE:3939370, mRNA, complete cds.
ACCESSION  BCO11711
VERSION    BCO11711.1      GI:15079815
KEYWORDS   MGC.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1592)
Strausberg, R.
Direct Submission
Submitted (30-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nigrl.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Boufard, G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantrijop, S., Thomas, P.J.,
Tjongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNI at: <http://image.ilni.gov>
Series: IRAL Plate: 27 Row: J Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7021050.

FEATURES

source

1. 1592

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:19518 IMAGE:3939370"

/tissue.type="Lung, small cell carcinoma"

/clone_lib="MIM MGC 7"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

460..1503

/codon_start=1

/product="unknown (protein for MGC:19518)"

/protein_id="AAH1711.1"

/db_xref="gi:15079816"

/translation="MKAKKMKKIGDMRFGKHKRDKIEKTKIKIQSEPTSEER
IRMKQERIQAKTRERKQAKERDYAEIOPDRHFGDDDELMTGVSYSYSMALN
ARPOSEKHEMDALYAQVKKPRNSPSPVDSNRSTSHNDRIORROFOCAKQDND
VEDRRRTYSEFOQPNAPATOGSRHSVYQVOMORQDERSSQAOQRTSSLFRO
SRKNASSVSDSWENYSPGEFOSAKENRYSYSGSRNGYLGGHFNARVWLETOE
ILROEORREOKKOPSEKPSNDSYKKVQDPSPAPKGPFPDOPSPSPQVAFIN
RLQTPKGRPFYS"

CDS

BASE COUNT 517 a 353 c 439 g 281 t
ORIGIN

alignment_scores:

Quality: 497.00 Length: 497
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x BC011711 ..

Align seg 1/1 to: BC011711 from: 1 to: 1592

860 GluThrLysLeuAsnThrValAspAspGlnLysAlaGlySerProSerar 876
|||||
10 GAGACTAAACTCAATACAGTGCATGACCAAGAAAGAGTTCTCCACAGCAG 59
876 gaAsValAlGlyProSerLeuGlyLeuLysLysSerSerSerLeuGluSerL 893
|||||
60 AGATGTGGTCTCCCTGGGCTCTGAAGAAAGTCAAGCTCATTTGAGAGATC 109

893 euglnthrAlaValAlaGluValThrLeuAsnGlyAspIleProPheHis 909
|||||
110 TGCAGACCGCAGATTGCCAGAGTACTTTGAATGGGATATCTCTTCCAT 159
910 ArProArProArGlyIleIleArgGlyArgGlyCysAsnGluSerPheAr 926
|||||
160 CGTCCACGGCCCGCATATCATCAGAGCGAGGATGCATATGAGACTTCAG 209
926 gAlaAlaIleAspLysSerTyrrAspLysProAlaValAspAspAspG 943
210 AGCTGCATCGCAAAATCTATGATTAACCCCGGATGATGATGATGATG 259
943 LuGlyMetGluThrLeuGluGluLysPheGluGluGluSerSerArgSerGly 959
260 AAGGATGAGACCTTGGACAGACACACAGAGAAATTCATAGATCAGGG 309
960 ArgGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGluAr 976
310 AGAGAGTGTGTATCCACAGCCAGATGATCAGCTCCACCTCTGGAGAG 359
976 gGlnMetAsnGlyAsnGlnGlnGlyLysGlyAspLysThrAspArgLysLysA 993
360 ACAATGATGGAACCAAGCAAGAAAGTGATTAAGACTGATGAAAGG 409
993 sPLysThrGlyLysGluLysLysLysAspArgAspLysGluLysAspLys 1009
410 ATTAAGATCGAAAGAAAGAAAGAAAGATAGATTAAGAGAAAGATTA 459
1010 MetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgPh 1026
460 ATGAAGCCCAAGAGAGAGTCTGAAGGCTTTGGAGACATCTTCAGGTT 509
1026 eGlyLysHisArgLysAspAspLysIleGluLysThrGlyLysIleLysI 1043
510 TGGCAAAACATCGAAAGATGACATTCAGAAACCGGTAAATTAATTA 559
1043 leGlnLysSerPheThrSerGluGluGluGluGluGluGluGluGlu 1059
560 TACAGGAATCTTTACATCAGAAAGAGAGAGATGAGATGAGAGAGAG 609
1060 GlnGluArgIleGlnAlaLysThrArgGluPheArgGluArgGlnAlaAr 1076
610 CAGGAGAGATTCACCAACCAACTCGAATTTAGGAGAGCAGACGCTCG 659
1076 gGluArgAspTyrrAlaGluIleGlnAspPheHisArgThrPheGlyCysA 1093
660 AGAGCGTGACTATGCTGAATTCAGATTTTCATCGACATTTGGCTGTG 709
1093 spAspGluLeuMetTyrrGlyGlyValSerSerTyrrGluGlySerMetAla 1109
710 ATGATGAGTTAATGATATGGGAGGATTTCTTCTATGAAGTTCCATGGCT 759
1110 LeuAsnAlaArgProGlnSerProArgGluGlyHisMetLeuAspAlaLe 1126
760 CTCAGACCTAGACCTCAGACCCACGAGAGGCGCATATGATGATCTT 809
1126 uTyrrAlaGlnValLysLysProArgAsnSerLysProSerProValAspS 1143
810 GTATGCCCAAGTCAAGAGCGCGGATTCCAAACTCCATCCTGAGACA 859
1143 eAsnArgSerThrProSerAsnHisAspArgIleGlnArgLeuArgGln 1159
860 GTACAGATCAACCTCTAGCAATCATGATGAGATACAGCTCTGAGGCAA 909
1160 GluPheGlnGlnAlaLysGlnAspGluAspValGluAspArgArgArgTh 1176
910 GAATTTCAAGCAAGCAAGCAAGATGAGATGATGAAGATGCTGCGCGGAC 959
1176 rTyrrSerPheGluGlnProTrrProAsnAlaArgProAlaThrGlnSerG 1193
960 CTATAGTTTGAACACCTGCGGAGACGAGCGGCGGCGGAGAGAGCG 1009


```

|||||
427 CTCTGGATGATGTAAGCTCTAGAAAGTCCCAACGATGAGGGCT 476
284 LeuGlyIleHsValValProPheSerAlaArgGlyArgThrLeuG1 300
|||||
477 CTGGGATTCCTAGTACGCCCTTCAGTGCCTGAGCGGAGAACCTGG 526
300 yLeuLeuValLysArgLeuGlyGlyLysAlaGluHsGluSln 317
527 GTATTAGTAAACGATTGAGAAAGGTGTAAGCTGAACATGAAATC 576
317 eupheaArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 333
577 TTTTTCGAGAAATGATTGATTCAGGATTAAATGATGCGACCTTCGA 626
334 AsnArgArgPheGluGlnAlaGlnHsMepHearGlnAlaMetArgTh 350
627 AATGAAAGATTGAAACAAACCAATATGTTTCCCAAGCCATGCGTAC 676
350 rProIleIleTrpPheHsValValProAlaAlaAsnLysGluInTyrg 367
677 ACCCATATTGTTGCTCCATGTGTTCCGACAAATTAAGACAGTATG 726
367 LuginLeuSerGlnSerGluLysAsnAsnTyrrTyrrSerSerArgPheSer 383
727 AACAACTATCCCAAGTGAAGAAACAATTAATCTTCAAGCCGTTTAAAC 776
384 ProAspSerGlnTyrrIleAspAsnArgSerValAsnSerAlaGlyLeuH1 400
777 CCTGACGCCAGTATATTGACACAGAGTGTGAACAGTGCAGGGCTTCA 826
400 sThValGlnArgAlaProArgLeuAsnHsProProGluGlnIleAspS 417
827 CACGCTGCAGAGACACCCCACTGAACACCCCGCTGACGACGATAGACT 876
417 eHsIleSerArgLeuProHsSerAlaHsProSerGlyLysProSer 433
877 CTCACCTACAGACTCTCATAGCGACACCCCTGGGAAAACCCACATCC 926
434 AlaProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerG1 450
927 GCTCAGGCTCGGACCTCAGACATGTATTAGTACGACTGTAAAGCAGTGG 976
450 yTrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
977 TTATATACACCAAAAAATAGGCAAGAGGCTTAATATCCAGCTTAAGAAAG 1026
467 LyrThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 483
1027 GTACGAAAGGTTTGGGATTACGACTCTCCAGAGATGTAAACAATAGGT 1076
484 GlySerAlaProIleTyrrValLysAsnIleLeuProArgGlyAlaAla11 500
1077 GGCTAGGCTCCACATCTATGTAAAAACAATCTCCCGGGGGGGCCCAT 1126
500 eGlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyV 517
1127 TCAGATATGCCGACTTAAGCAGACGACGACTTATAGAGTAAATGAGNG 1176
517 AlaAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 533
1177 TAGATTGTGTGGCAAAATCCCAAGAGAAAGTGTTCCTGCTGTGAAGAC 1226
534 ThrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAl 550
1227 ACCAAGATGGAAGAACTGTGAGCTTCTGCTTCCGCCAGAGAGACC 1276
550 aPheHsProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProL 567
1277 CTTCCACCCAAAGGAATGAATGACAGCCAAAGCAGATGACGATTCCAA 1326
567 yGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThr 583
|||||

```

```

1327 AAGAAAGCAAAAGCAGAAAGATGAGATATTGTTCTTACACCTGATGGCAC 1376
584 ArgGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaG 600
1377 AGGGAATTTTGCACATTTGAAAGTCCCACTTAGT.GATTACAGATCTGAG 1425
600 LysLeuGlyValSerValLysGlyAsnArgSerLysGluAsnHsAlaAsp 616
1426 GCCTTGGTGTGCTGCTCAAAAGTAAACCGTCAAAAGAGAACACGACGAT 1475
617 LeuGlyIlePheValLysSerIleIleAsnGlyLysAlaAlaSerLysAs 633
1476 TTGGGAATCTTGTCAAGTCCATTATTAATGAGAGACGACATCTTAAGA 1525
633 pGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerL 650
1526 TGGAAGGCTTGGGGATGATGATCAACTGATGACGTAATGAGATATCC 1575
650 euleuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMet 666
1576 TGTGGGCAAGCAAAACCAAGATGCCATGGAACCTTAAGAGGTCTATG 1625
667 SerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgAr 683
1626 TCTACTGAAGCCAAATAACGAGGAATGATCCAGCTTATTTGCAAGAG 1675
683 gIleSerLysCysAsnGluLeuLysSerProGlySerProProGlyProg 700
1676 AATAAGCAAGTGCATAGTGAAGTCACTGGGAGCCCTTGACCTG 1725
700 LuLeuProIleGluThrAlaLeuAspAspArgGluArgIleSerHis 716
1726 AGCTGCCATTGAAACACCGTTGATGTAGAAACGAAGATTTCCCAT 1775
717 SerLeuTyrrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAl 733
1776 TCCCTCTACAGTGGGATTTGAGGGCTTGTATGATATGCCAGCAAAATGC 1825
733 AlaLeuSerArgIleMetGly 740
1826 TGCCCTCAGTAGGATTAATGGGT 1847

seq_name: gb_pr:AK027735
seq_documentation_block:
LOCUS AK027735 2718 bp mRNA linear PRI 15-MAY-2001
DEFINITION Homo sapiens cDNA FLJ14829 fis. clone OVARC100945, moderately
similar to Rattus norvegicus mRNA for atypical PKC specific binding
protein.
ACCESSION AK027735
VERSION AK027735.1 GI:14042633
KEYWORDS oligo cloning; fis (full insert sequence).
SOURCE Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_11b:OVARC1
clone:OVARC100945.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE
AUTHORS Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Mishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Matsushima,M., Hosotani,T., Kaku,K., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakanura,Y., Nagahari,K., Masuhara,Y., Ninomiya,K. and Iwanaghi,T.
MEDO human cDNA sequencing project
TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2718)
AUTHORS Isegai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Takao Isegai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

```

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan. cDNA full insert sequencing. Research Association for Biotechnology: cDNA library construction. Helix 5'-3'-end one pass sequencing and cDNA selection. Research Institute (supported by Japan Key Technology Center etc., and Department of Virology, Institute of Medical Science, University of Tokyo).

FEATURES

```
source
1. "tissue"="Homo sapiens"
/organism="taxon:9606"
/ID="OVARC1000945"
/tissue.type="ovary, tumor tissue"
/clone.lib="OVARC1"
/note="cloning vector: pME18SFL3"
33. 2630
/note="unnamed protein product"
```

CDS

	LABOUR DATA	
BASE COUNT	840 a	586 c
ORIGIN	688 g	604 t

alignment_scores:	508
quality:	420.00
ratio:	0.830
Percent Similarity:	99.606
	Percent Identity:
	99.606
	gaps:
	1

alignment_block: AK027735

Align seg 1/1 to: AK027735 from: 1 to: 2718

234 GILUYSGLINGLUGINASPGLIAPSGLYTHRTGLUGIAPSPASPSERATG 250
327 GAGAAACAGAAACAGGATGAGAGATGGGACAGAGAGAGATTAACAGTCGT 376
250 IGLUPROVALGILYHSAIAAPRHTGLYLEUGLHLSILEPROAPHPES 267
377 TGAACCTGTTGGACATGCTGACAGCGGCTTGGAGCATATACCACTTTT 426
267 ERLEUASPAERMEVALIYSLSEUVALIGLUALIPROASNAPGLYGLYPRO 283
427 CTGTGATGATATGGTAAAGCTCGTAGAAGCTCCCAACGATGGAGGGCT 476
284 LEUGLYILEHSAVALIPROPHESERALAARGLYLYRTHRTLEUGL 300
477 CTGGGAAATCCATGTAGTACGCCCTTCAGTGTGTCAGGCGGCAAACTGGG 526
300 YLEUENVALIYARGLEUULYSGLYGLYLSALAGLHLSGLUASNIL 317
527 GTTATTAGTAAACAGATTGGAGAGAGTGGTAACCTCAACATGTAAATC 576
317 EUPHEARGLUASNAPSYLLEVALARTGLEASNAPGLYASPLENARG 333
577 TTTTTCGTAGAAATATTTGCATCTGTCAGGATTAATGATGAGCAACTTCGA 626

[illegible]

```

633 pGlyArgLeuArGValAsnAspGlnLeuIleAlaValAsnGlyGluSerL 650
|||||
1526 TGGAGGCGCTTCGCGTGAATGATCACTGATGCGCAATTAATGAGCAATCCC 1575
|||||
650 euleuGlyLysThrAsnGlnAspAlaMetGluThrIleArgArgSerMet 666
|||||
1576 TGTGGGGAGACAAACCAAGATGCGCATGGAACCCCTAAGAGGCTATG 1625
|||||
667 SerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleAlaValArg 683
|||||
1526 TCTACTGAGGCATTAACAGACGATGATCCAGCTTATGTTGCAAGAG 1675
|||||
683 gLleSerLysCysAsnGluLeuLysSerProGlySerProGlyProG 700
|||||
1676 AATAGGAGAGGCAATGAGCTGAGTCACTGAGGCCCTGAGACTG 1725
|||||
700 lueuProlIleGluThrAlaLeuAspAspArgGluArgIleSerHis 716
|||||
1726 AGCTGCCATTTGAAACAGCGTTGATGATGAGAACGAGAAATTTCCCAT 1775
|||||
717 SerLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsn 733
|||||
1776 TCCTCTACAGATGGGATGAGGGCTTGATGATCGCCACAGAAATGC 1825
|||||
733 aAlaLeuSerArgIleMetGly 740
|||||
1826 TGCCCTCAGTAGATGAATGGCT 1847
|||||
seq_name: gb_pr:AK000761

seq_documentation_block:
LOCUS AK000761 2696 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens CDNA FLJ20754 fis, clone HEP02246.
ACCESSION AK000761
VERSION AK000761.1 GI:7021050
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HePg2 CDNA to mRNA, clone_1lb:HEP
clone:HEP02246.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigetani,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
TITLE 2 (bases 1 to 2696)
REFERENCE Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
AUTHORS Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1. 2696
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP02246"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_1lb="HEP"
/feature="cloning vector pME18SFL3"
45..1088
CDS

```

```

/note="unnamed protein product"
/codon_start=1
/protein_id="BA91366.1"
/db_xref="GI:7021051"
/translation="MKAKGMKKGIDMRFGRKRDQKIEKTKIKIOESTSEER
IRKROERIOATREFRERORAREYAEIÖDHRTFGCDELMGTGVSSTEGSMALN
ARKPOSREGIMDALYAQVKPKNRSPSPVDSNRSPSHNRIORLROEFOAKQOED
VEDRRRTYSFEPWPNAARPATOSGRHSVSVEYOMOROEERESODAOROYSSLPRO
SRKNASSVSODSWEONYSPGEGFOSAKENRYSYSGSRNGYLGCGFNARVMLETOE
LROBORRKKDQMKQPPSGSPSNYSYKKYQDPSVAPPKGPPODVPSPSPQVARTLN
RIQTPKGRPRFS"
BASE COUNT 703 a 611 c 695 g 687 t
ORIGIN

alignment_scores:
Quality: 356.00 Length: 356
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-09-757-781-2 x AK000761
Align seg 1/1 to: AK000761 from: 1 to: 2696

1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysGlyMetLe 1017
|||||
18 AAGATAGAGATTAAGAGAGAGATTAATGAAGCAAGCAAGGAGATGCT 67
|||||
1017 ULysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAsp 1034
|||||
68 GAAGGCTTGGAGACATGTCAGCTTGGCAACATGCAAAAGATGACA 117
|||||
1034 YSILeGluLysThrGlyLysIleLysIleGlnGluSerPheThrSer 1050
|||||
118 AGATTGAGAAACCGGTAATTAATAATACAGAAATCTTTCATCAGAA 167
|||||
1051 GluGluArgIleArgMetLysGlnGlnGluArgIleGlnAlaLysTh 1067
|||||
168 GAGGAGAGATACGAATGAGAGAGAGAGAGAGAGATTCACCAAGAAC 217
|||||
1067 PArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGlu 1084
|||||
218 TCGAATAATTAGGAGACACAGCTCGAGAGCGTACATGCTGTAATTC 267
|||||
1084 LAspPheHisArgThrPheGlyCysAspAspGluLeuMetTyrGly 1100
|||||
268 AAGATTTTCATCGGACATTTGGCTGTGATGATGATTAATGATGGGGA 317
|||||
1101 ValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSer 1117
|||||
318 GTTCTCTCTAATGAAGGTTCATGCTCAACGCTAGACCTCAAGAGCC 367
|||||
1117 CArgGluGlyHisMetLeuAspAlaLeuTyrAlaGlnValLysLysPro 1134
|||||
368 ACGAATAAGGCAATAGATGATGCTTGTATGCCCCAATCAAGAGCCGC 417
|||||
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
|||||
418 GGAATTCGAACCCCTCACCTGTAGACATTAACAGATCACTCTTACGAA 467
|||||
1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGln 1167
|||||
468 CATGATCGGATACAGCGCTGAGGCAAGAAATTCGCAAGCAAGCAAGA 517
|||||
1167 pGluAspValGluAspArgArgTyrTyrSerPheGluGlnProTyrP 1184
|||||
518 TGAAGATGAGAGATCGTGGCGGACCTATATGTTTACGCAACCCCTGC 567
|||||
1184 rOAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
|||||
568 CGAAGCAGGCGGCGAGCAGAGAGCGGCGACATCGTGTCCGTGGAG 617
|||||

```



```

53 TCACAACGAGAGTGTGAACAGTGCAGGCTTCACACCGTCGACAGACAC 102
406 roArgLeuAnHisProProGluGlnIleAspSerHisSerArgLeuPro 422
103 CCCGAGTGAACCCAGCCGCTGACAGAGATGAGCTCTCCTCACTAAGATACCT 152
423 HisSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaPr 439
153 CATACCGACACACCCCTGGGAAAACACATCCGCTCCAGCCTGGGACCC 202
439 GlnHisValPheSerThrValSerSerGlyTyrAsnThrLysLysI 456
203 TCACAAGTATATTAGTACGACTGATGACGAGTGGTGTATTAACACCAAAAA 252
456 legLysArgLeuAnHisIleGlnLeuLysLysGlyThrGluGlyLeu 472
253 TAGGCAAGAGGCTTAAATCCACAGCTTAAGAAAGGTACAGAAAGTTTGGGA 302
473 PheSerIleThrSerArgAspValThrIleGlySerAlaProIleTyr 489
303 TTCAGATCACTTCCAGAGATGTAAACATAGGTGGCTAGCTCCAACTTA 352
489 rValLysAsnIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeu 506
353 TGTGAAAAACATTTCTCCCGGGGGGGCCGCTCAGAGATGCCGACTTA 402
506 yAlaGlyAspArgLeuIleGlnValAsnGlyValAspLeuValGlyLys 522
403 AGGCGAGGAGACAGACTTATAGAGTAAATGAGATGATTTAGTGGGCAAA 452
523 SerGlnGluGluValValSerLeuLeuArgSerThrLysMetGluGly 539
453 TCCCAAGAGGAAGTTGTTTCTGCTGTGAGAAAGCACCAAGATGAGAAAG 502
539 rValSerLeuLeuValPheArgGlnGluAspAlaPheHisProArgGlu 556
503 TGTGAGCTTCTGTGCTTTTCCGAGAGAACAGCCCTTCCACCCAGAGGAC 552
556 eulaAsnAlaGluProSerGlnMetGlnIleProLysGluThrLysAlaGlu 572
553 TGAATGAGAGCCACGACGATCCAGATTCACAAAAGAAACGAAAGACAA 602
573 AspGluAspIleValLeuThrProAspGlyThrArgLysPheLeuThr 589
603 GATGAGAGATATGCTTCTTACACCTGATGACACGAGGGAATTTCTGACATT 652
589 eGluValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerVal 606
653 TGAAGTCCCACTTATGATTCAGATCTGCAAGGCTGTGTCTAGTCTCA 702
606 yGlyAsnArgSerLysGlnAsnHisAlaAspLeuGlyIlePheValLys 622
703 AAGGTAAACGGTCAAAAAGAGAACCCAGCATCTGGGAATCTTTGTCAAG 752
623 SerIleIle 625
753 TCCATTTATT 761
seq_name: gb_est2:BE792557
seq_documentation_block:
LOCUS BE792557 725 bp mRNA linear EST 20-SEP-2000
DEFINITION 60158334.1P1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3939370 5',
ACCESSION BE792557
VERSION BE792557.1 GI:10213755
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgs.ncl.nih.gov/.

```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at: Image.lnl.gov
Plate: ILGM/88 row: b column: 11
High quality sequence stop: 725.

FEATURES
source Location/Qualifiers
1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3939370"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOT87; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCGAGAG(c). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 256 a 118 c 207 g 144 t
ORIGIN

alignment_scores:
Quality: 225.00 Length: 225
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x BE792557 ..

Align seg 1/1 to: BE792557 from: 1 to: 725

```

867 AspAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuG 883
3 GATGACCGAGAAAGCAAGTTCCTCCAGAGATGATGGGCTCTCCCTCGG 52
883 yLeuLysLysSerSerSerLeuGluSerLeuGlnThrAlaValAlaGlu 900
53 TCGAAGAGAGTCAACCTCATTTGAGAGTCTCCAGACCGCAGTTGCCAG 102
900 aThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIle 916
103 TGAATTTGATGGGCAATTTCTTCCATCTCCACGCGCGCGGATATTC 152
917 ArgGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyr 933
153 AGAGGAGAGGGATGCAATGAGACTTCAGAGCTGCCATCGACAAATCTTA 202
933 rAspLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluG 950
203 TGAATAAACCCCGGTAGATGATGATGATGATGATGATGATGATGATGAT 252
950 LuAspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAla 966
253 AAGACACAGAGAAAGTAAGTTCAGATCAGGAGAGAGAGTCTGATCCACGC 302
967 SerAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnG 983
303 AGTGATCAGCCTTCCCACTCTCTGGAGAGACAAATGATGAAACCAAGA 352
983 uLysGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysL 1000
|||||

```

353 GAAAGCGATGAAGCTGATGAAAAAGATTAAGTGAAGAAAGA 402
 1000 ysltyspataasplysglulysaspmtlysatalyslysglymet 1016
 403 AGAAGATTAGATTAGAGACAGATTAATGAAGCCAGAAAGGATG 452
 1017 leulysglyleugllyaspmetpheargpheglylyshisarglysaspas 1033
 453 CTGAAGGGCTTGGGACATGTCAGGTTGGCAACATCGAAAAAGATGA 502
 1033 plysllleugllythrgllystlelyslleaglulserptherberg 1050
 503 CAAGATTGAAGAAACGGTAAATAAATAACAGAAATCCTTACATCAG 552
 1050 luclugluarglleargmetlysinglulugluarglleaglinalys 1066
 553 AAGAGGAGAGATTACGATGACAGAGACAGAGAGATTCAAGCCAAA 602
 1067 thtarggluaphearggluarglinalaarggluarasptyrallagluil 1083
 603 ACTCGAATTTAGGAAACAGACAGCTCGAGAGCGTACTATGCTGAAT 652
 1083 eglinaasphelisargthrphegly 1091
 653 TCAAGATTTCATCGACATTTGGT 677
 seq_name: gb_est2:BM466811

seq_documentation_block:

LOCUS BM466811 1150 bp mRNA linear EST 05-FEB-2002
 DEFINITION AGENCOURT_6431318 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5503072
 5', mRNA sequence.

ACCESSION BM466811
 VERSION BM466811.1 GI:18515853
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

TITLE NIH-MGC http://mgs.nci.nih.gov/

AUTHORS 1 (bases 1 to 1150)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov

COMMENT Tissue Procurement: ATCC

COMMENT cDNA Library Preparation: Life Technologies, Inc.

COMMENT DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

COMMENT Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM12142 row: d column: 17

High quality sequence start: 9

High quality sequence stop: 652.

Location/Qualifiers

1. 1150

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5503072"

/clone_1lb="NIH_MGC_67"

/tissue_type="retinoblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pCMV-Sport6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT"

Average insert size 1.75 kb. Library constructed by Life

Technologies. 309 g

BASE COUNT 294 a 310 c 235 t 2 others

ORIGIN

alignment_scores:

Quality: 213.00

Length: 213

Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-757-781-2 x BM466811 ..

Align seq 1/1 to: BM466811 from: 1 to: 1150

1112 AlAargProginserProAargGluGlyHismetelsapalaleuTyraI 1128
 39 GCTAGACCTCGAGACCCAGGAAAGGCGCATATGATGATGCTTTGTATGC 88
 1128 aglnvallylsyProAargAsnSerlyspSerProvalAspserAsna 1145
 89 TCAAGTCAGAAAGCGCGGAATTCGAAACCTCACTACCTAGACAGTACA 138
 1145 rgSerThrProSerAsnHisasparglllelnaagleuAarglulphe 1161
 139 GATCAACTCCTAGCAATCATGATCGATACAGCGCTGAGCCAGAAATTT 188
 1162 glnglnalalysglnaspgluaspyalglulaspargargthrtyrse 1178
 189 CAGCAGCAAGCAAGATGAGATGATGATGATGATGATGATGATGATG 238
 1178 rPheGluGlnProTTPProAsnAlaargProAlaThrGlnSerGlyargH 1195
 239 TTTGAGCAACCTGGCGGAGCGAGCGCGCGGCGAGCGAGCGCGGCGAC 288
 1195 lsservalserValglulValglmetGlnaargGlnaargGlnGluarg 1211
 289 ACTCGGTGTCGTGGAGGTGAGATGAGATGAGATGAGATGAGATGAG 338
 1212 gluserSerGlnGlnalaglnaargGlnaargGlnaargGlnaargGln 1228
 339 GAGAGCTCCAGCAGAGCCAGCCAGCCAGCTTCTGCTTCCGCAAG 388
 1228 rArglysaAnaIaservalSerGlnaspSerTrpGlnaasnTyrS 1245
 389 CAGGAAAAATGCCAGCTCGTCTCCAGACTCTTGAGAGCAAGAACTACT 438
 1245 erProGlyGlyGlyPheGlnSerAlalyGlnaasnProAargTyrSer 1261
 439 CCCCTGGGGAAGGCTTCCAGAGTGCAGAAAGAACCCAGGTACTCCAG 488
 1262 TyrGlnGlySerArgAsnGlyTyrleuGlyHysGlyPheAsnAlaAr 1278
 489 TACCAAGCTCCAGGACGCTACCTGGAGAGACATGCTTCAAGCCAG 538
 1278 gValmetleuGlnurthrglnGlnleuAargGlnGlnaargarglysg 1295
 539 GGTCAATGCTGGAACCTCAGAGCTCTTCCGAGAGAACAGAGCGGAAG 588
 1295 luGlnGlnmetlylsyGlnProProSerGlyGlyProSerAsnTyrasp 1311
 589 AGCAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
 1312 SerTyrlylsyValGlnaspProserTyrAlaProPro 1324
 639 TCGATTAAGAAAGTCCAGAGACCCAGTTACGCCCTCTCC 677
 seq_name: gb_est1:AM780417

seq_documentation_block:

LOCUS AM780417 599 bp mRNA linear EST 12-MAY-2000
 DEFINITION ho2l808.x1 NCI CGAP_C014 Homo sapiens cDNA clone IMAGE:3038007 3'
 similar to TR:092340 Q92340 AYPICAL PKC SPECIFIC BINDING PROTEIN.
 ;, mRNA sequence.

ACCESSION AM780417
 VERSION AM780417.1 GI:7795020
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Thu Jul 25 08:38:28 2002

us-09-757-781-2.0lig.rst

Page 4

REFERENCE 1 (bases 1 to 599)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL, send email to:
info@image.lnl.gov
Possible reversed clone: polyA not found
Seq primer: -400P from GIBCO
High quality sequence stop: 412.
Location/Qualifiers
1..599
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3038007"
/clone_lib="NCI-CGAP Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.7 Kb. Life technologies catalog #:
11531-019"
BASE COUNT 130 a 157 c 117 g 195 t
ORIGIN
alignment_scores:
Quality: 189.00 Length: 189
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent identity: 100.000
alignment_block:
US-09-757-781-2 x AW780417/rev ..
Align seg 1/1 to reverse of: AW780417 from: 1 to: 599
808 leuserProaspValaspProvalaleualaphenlathargluclyphgl 824
|||||
567 TTGAGTCAGATGTGATCCAGTCTTCACAGAGAGATTGG 518
824 YARGInserMetserglulysArgThrlysglnPheserAspAlaSerG 841
|||||
517 ACGTCAGAGTATGTCAGAAAACCCACAAACCAATTTTCAGATGCCAGTC 468
841 InleuAspPheVallyThrArglysserlysserMetAspLeuGlyIle 857
467 AATTGATTTCGTTAAACAGAAAATCAAAACCAATGATTAGGATA 418
858 AlaAspGluThrlyleuAsnThrValAspAspGlnlyAspIleSerP 874
|||||
417 GGTGACAGACATTAACATACAGTGTGATGACCAAGAAAGAGTTCTCC 368
874 oSerArgAspValGlyProSerleuglyleuLylysserSerleug 891
367 CACGACAGATGTGGCTTCCTCGGTCTGAAAGATCAAGCTCATTTGG 318
891 InSerleugInThrAlaValAlaGluValThrleuAsnGlyAspIlePro 907
|||||
317 AGATCTGCAGACCGCAGTGGCCGAGGTACTTGAATGGGATATTCCT 268
908 PheHisArgProArgProArgIleIleArgGlyArgGlyCysAsnGlu 924
267 TTCATGTCTCAGCGCCGCGATATACAGAGCGAGGATGCAATGAGAG 218

924 rPheArgAlaAlaIleAspIlysserlyrAspIlyProAlaValAspAsp 941
|||||
217 CTCAGAGCTCCCATCGACAAATCTTATGATTAACCCGGGCTGATGARG 168
941 SPASpGluGlyMetGluThrleuGluGluAspThrGluGluSerArg 957
|||||
167 ATGATGAAGAGCATGGAGACCTGTGAGAAAGACACAGAAAGAACTTCAAG 118
958 SerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSerle 974
117 TCAGGAGAGAGATCTGTATCCACAGCCGATATCAGCTCCCACTTCCT 68
974 uGluArgGlnMetAsnGlyAsnGlnGluLylysglyAspIlyThrAspArg 991
67 GGAGAGACAAATGATGAGAAACCAAGAGAAAGTGATTAAGCTATGAA 18
991 ysllyAspIlyThrGly 996
17 AAAGGATTAACCTGGA 1
seq_name: gb_est2:BE693704
seq_documentation_block:
LOCUS BE693704 645 bp mRNA linear EST 11-SEP-2000
DEFINITION RC0-BT0387-100800-013-f05 BT0387 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE693704
VERSION BE693704.1 GI:10080864
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 645)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
20202663
CONTACT: Simpson A.J.C.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=62=RC0-BT0387-100
800-013-f05&f3=2000-08-10&f4=1)
800-013-f05&f3=2000-08-10&f4=1)
Seg primer: puc 18 forward
High quality sequence start: 45
High quality sequence stop: 644.
Location/Qualifiers
1..645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0387"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site: 1: SmaI; Site: 2:
SmaI; A mini-library was made by cloning products derived
from ORFESTS PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 122 a 176 c 115 g 232 t
ORIGIN

alignment_scores:
 Quality: 188.00 Length: 188
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-757-781-2 x BE693704/rev ..

Align seg 1/1 to reverse of: BE693704 from: 1 to: 645

```

838 ASPALASerGlnLeuAspPheValLysThrArgLysSerIysSerMetas 854
|||||
610 GATGCCAGTCATATGGATTTCGTTAAACGAAATCAAAAGCATGAGA 561
|||||
854 pLauGlyIleAlaAspGluThrLysLeuAnThrValAspAspGlnLysA 871
|||||
560 TTTAGGATATAGCTGACGACACTAACTCAATACAGTGAGACCAAG 511
|||||
871 laGlySerProSerArgAspValGlyProSerLeuGlyLeuLysLysSer 887
|||||
510 CAGGTTCTCCACAGCAGAGATGGGTCTTCCCTGGGTCTGAGAGACTCA 461
|||||
888 SerSerLeuGlnSerLeuGlnThrAlaValAlaGluValThrLeuAnG 904
|||||
460 AGCTCATTCGAGAGTCTGCAGACCGCAGTTGCCGAGGTGACTTGAATG 411
|||||
904 YASpIleProPheHisArgProArgProArgIleIleArgGlyArgGly 921
|||||
410 GGATATTCCTTCCATCCTCCACGCGCGGATATACAGAGCGGCGAT 361
|||||
921 YASAnGlnSerPheArgAlaAlaIleAspLysSerIysAspLysProAla 937
|||||
360 GCATGTGAGACTTCAGAGCTGCCATCGACAATTTATGATTAACCCGCG 311
|||||
938 ValAspAspAspAspGluGlyMetGluThrLeuGlnGluAspThrGlu 954
|||||
310 GTAGATGATGATGATGAAGCATGGAACCTTGGAAGAAAGACAGAGA 261
|||||
954 userSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnPro 971
|||||
260 AAGTTCAGATTCAGGAGAGAGTCTGTATCCACAGCCAGATGATCAGCTT 211
|||||
971 erHisSerLeuGlnArgGlnMetAnGlnYASnGlnGlnLysGlyAspLys 987
|||||
210 CCACACTCTCGAGAGACAAATGATGAAACCAAGAAAGGATGATAG 161
|||||
988 ThrAspArgLysLysAspLysThrGlyLysGluLysLysLysAspArg 1004
|||||
160 ACTGATAGAAAAAGATAAACCTGAAAAAGAAAGAAAGAAAGATAGAGA 111
|||||
1004 pLysGluLysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeu 1021
|||||
110 TTAGGAGAAAGATTAATGAAGCAAGAAAGGAATGCTGAGGCGTTGG 61
|||||
1021 LysAspMetPheArg 1025
|||||
60 GAGCAGATGTCAGG 47

```

seq_name: gb_est2:BG910244

seq_documentation_block:

LOCUS BG910244 694 bp mRNA linear EST 05-JUN-2001
 DEFINITION 602805735F1 NCI CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4937865
 5' mRNA sequence.

ACCESSION BG910244
 VERSION BG910244.1 GI:14290720

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 694)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L14M10872 row: f column: 10
 High quality sequence stop: 694.

FEATURES
 source
 Location/Qualifiers
 1..694
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4937865"
 /clone_lib="NCI CGAP_Brn67"
 /issue_type="anaplastic oligodendroglioma with 1p/19q
 loss"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Brain; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.3 kb. Constructed by Life
 Technologies. Note: this is a NCI CGAP Library."

BASE COUNT 186 a 189 c 205 g 114 t
 ORIGIN

alignment_scores:
 Quality: 179.00 Length: 193
 Ratio: 0.932 Gaps: 1
 Percent Similarity: 99.462 Percent Identity: 99.482

alignment_block:
 US-09-757-781-2 x BG910244 ..

Align seg 1/1 to: BG910244 from: 1 to: 694

```

1108 MetAlaLeuAsnAlaArgProGlnSerProArgGluGlnHisMetMetas 1124
|||||
43 ATGGCTTCAACGCTAGACCTCAGAGCCACAGAAAGGATATGATGGA 92
|||||
1124 pAlaLeuTyrlaGlnValLysLysProArgAnSerIysProSerProy 1141
|||||
93 TGCCTTGTATGCCCAAGTCAAGAAACCCGCGAATTCAAACCTTCACCTG 142
|||||
1141 alaSerSerAsnArgSerThrProSerAsnHisAspArgIleGlnArgLeu 1157
|||||
143 TAGACAGTACAGATCAACTCTTAGCAATCATGATGATACAGGCTCTG 192
|||||
1158 ArgGlnGluPheGlnAlaLysGlnAspGluAspValGluAspArg 1174
|||||
193 AGCCAAAGATTTACAGCAAGCAAGCAAGATGAATGTAAGAAATGCTCG 242
|||||
1174 gArgThrTySerPheGlnGluGlnProTrpProAsnAlaArgProAlaThr 1191
|||||
243 GCGGACCTATAGTTTGTAGCAACCTTGCCGCAACCCAGCGCGGAGACG 292
|||||
1191 InSerGlyArgHisSerValSerValGluValAlaGlnArgGlnArg 1207
|||||
293 AGAGCGGGCGGCTCGGTCTCGTGGAGGTGCAGATGCAGGCGGAGCGG 342
|||||
1208 GlnGlnGluArgGluSerSerGlnGlnAlaGlnArgGlnTyrrSerSer 1224
|||||
343 CAGGAGGAGCGCGAGAGCTCCAGCAGGCGCCAGCCAGTACAGCTCTCT 392
|||||
1224 uProArgInSerArgLysAsnAlaSerSerValSerGlnAspSerTrp 1241
|||||
393 GCCTGCGCAAGCAGAAAAATGCCAGCTCGGTCTCCAGGAGCTCTTGGG 442

```

```
1241 luc1nasntyrserproglugluglypheclnseralalysclunapro 1257
|||||
443 acgacaaactactccctgggaaaggcttcacagatgcccaagaaacccc 492
|||||
1258 argtyrsersttyrglnclyserargasnlytyrleuclglyhlsgl 1274
|||||
493 aggtactccagctaccacagagctccagaaagcgtctggagagacatgg 542
|||||
1274 ypheasna1aargvalmetleu.gluThrGlnGluLeuLeuArgGlnGlu 1290
|||||
543 cttcacagcagcaggtcatgctgggaaacacagagctcctgcacgaa 592
|||||
1291 glna1aarglysglucglnclmetlys 1299
|||||
593 cagagcgcgaaagagcagcagatgaa 619

seq_name: gb_est2:B1093826

seq_documentation_block:
LOCUS B1093826 809 bp mRNA linear EST 20-JUN-2001
DEFINITION 60286060P1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:5001744 5',
mRNA sequence.
ACCESSION B1093826
VERSION B1093826.1 GI:14512156
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 809)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11035 row: 1 column: 01
High quality sequence stop: 759.
Location/Qualifiers
1..809
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site:1; NotI;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
FEATURES
source
BASE COUNT 252 a 174 c 191 g 192 t
ORIGIN
alignment_scores:
Quality: 164.00 Length: 164
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x B1093826 ..
Align seg 1/1 to: B1093826 from: 1 to: 809
577 valleurthproaspolythrarggluubheuthrphggluvalprole 593
|||||
```

```
57 gttcttacaccctgattgacccagcaggaatttctgacatttgagctccacct 106
|||||
593 uasnasperclyseralaglyleuGlyValSerValysGlyasnarg 610
|||||
107 taatgattcagagatctgcagagccttggtcagtgctcaaaagcttaaccgt 156
|||||
610 erlysgluasnhisalasprenuglyllephevalysSerlleasn 626
|||||
157 caaaagaaacaccagcagacatttgggaatcttggcnaagctcatattat 206
|||||
627 glyglyalalalaserlyaspglyargleuargyalaasnarglnleu 643
|||||
207 ggagagcagacatctpaaagatggaagccttcgggtgaatgatactgat 256
|||||
643 ealaya1asnuglyluserleuGlylysThransglnaspalametg 660
|||||
257 acgagtaaatggaatctcctgttggcnaagcnaaacnaatgctcagg 306
|||||
660 luthrleuargysermetserthrngluglyasnlyasarglymetlle 676
|||||
307 aaaccctaaagagctctatgctactgaagcgaatpaaacgaggaatgattc 356
|||||
677 glnleuilevalalalargarglleserlyscysasnuglyleuLysSerPr 693
|||||
357 cagcttattgttgcaagagaaatgaaagcagtgcaatgagctgaactcacc 406
|||||
693 oclyserproglglyprogluLeuProilegluthrleuaspasp 710
|||||
407 tggagaccccccctggacctgagactgcccattgaaacagcgttgatgata 456
|||||
710 rggluargarglleserhisserleutysercylglcuglyleuasp 726
|||||
457 gagaaagaaatgttccatctccctfacaatggagctgaaggccttgat 506
|||||
727 gluserproserargasnala1aleuserargylemetgly 740
|||||
507 gaatgccccaagcagaatctgcctccctcactagatpaaatgga 548

seq_name: gb_est2:B1870563

seq_documentation_block:
LOCUS B1870563 575 bp mRNA linear EST 11-OCT-2001
DEFINITION 60339404P1 NIH_MGC_90 Homo sapiens CDNA clone IMAGE:5404067 5',
mRNA sequence.
ACCESSION B1870563
VERSION B1870563.1 GI:16044236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 575)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12030 row: 9 column: 12
High quality sequence stop: 575.
Location/Qualifiers
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1b="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
FEATURES
source
```

/lab.host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 182 a 139 c 128 g 126 t
ORIGIN

alignment_scores:
Quality: 162.00 Length: 162
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x B1870563 ..

Align seg 1/1 to: B1870563 from: 1 to: 575

309 G1G1Y1YLSAAG1UHS1G1UASnLeuPheargG1UASnAspCysIleVa 325
|||||
12 GGGGTAAGCTGACATGAATAATCTTTTCGTGAGATGATTCATTTCT 61
325 1Arg1LeaSnAspG1YAsnLeuArgAsnArgArgPheG1UAG1nH 342
|||||
62 CAGGATTAATGATGGCGACCTTCGAAATAGACATTTGAACAACACCAAC 111
342 1sMetPheArgG1nAlaMetArgThrProIleIleThrPheHisVal 358
|||||
112 ATATGTTTCCCAAGCCATCGTACACCCATCATTTGGTTCATGTGTT 161
359 ProAla1AlaSnLysG1UAG1nTYrG1U1nLeuSerG1nSerG1U1YsAs 375
|||||
162 CCGGACGCAATATAAGAGCAGTATGACAACTATCCCAAGTGAGAA 211
375 nAsnTYrTYrSerSerArgPheSerProAspSerG1nTYr1LeaSpAsnA 392
|||||
212 CAATTACTATTCAAGCCGTTTTCAGCCCTGACAGCCACTATATTGACAACA 261
392 rGSerValAsnSerAlaG1YLeuHisThrValG1nArgAlaProArgLeu 408
|||||
262 GGGGTGTGAACAGTGCAGGCTTCACAGGTCGACAGAGCACCCGCACTG 311
409 AsnHisProProG1UAG1n1LeaSPSerHisSerArgLeuProHisSerAl 425
|||||
312 AACCAACCCGCTGAGCAGATAGACTCTCACACAGACTTACCTCATACGC 361
425 aHisProSerG1YUyProProSerAlaProAlaSerAlaProG1nAsnV 442
|||||
362 ACACCCCTCGGGAAAMCCACCATCCGCTCCAGCCTCGCACCTCAGATG 411
442 a1PheSerThrThrValSerSerG1YTYrAsnThrLysLys1LeG1Y1Ys 458
|||||
412 TATTTAGTACGACTGTAAAGCAGTGGTTATTAACACCAAAAAAATAGGCAAG 461
459 ArgLeuAsn1LeG1nLeuLysLysG1YThG1U1Y 470
|||||
462 AGGCTTAATATCCAGCTTAAGAAAGTACAGAAAGT 497

seq_name: gb_est2:BG473458

seq_documentation_block:

LOCUS BG473458 660 bp mRNA linear EST 21-MAR-2001
DEFINITION 602516255F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4648117 5',
mRNA sequence.

ACCESSION BG473458
VERSION BG473458.1 GI:13405733

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 660)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCM425 row: m column: 14
High quality sequence stop: 656.
Location/Qualifiers

FEATURES
source

1..660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4648117"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 204 a 135 c 172 g 149 t
ORIGIN

alignment_scores:
Quality: 159.00 Length: 159
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x BG473458 ..

Align seg 1/1 to: BG473458 from: 1 to: 660

570 LysA1AG1AspG1UAsp1LeValLeuThrProAspG1YThrArgG1uPh 586
|||||
88 AAACGACAGATGACGATATGTTCTTACACTGATGACACGAGGATTT 137
586 eLeuThrPheG1UValProLeuAsnAspSerG1YSerAlaG1YLeuG1Y 603
|||||
138 TCTGACATTTGAAGTCCACCTTAATGATTCAGAGATCTCAGGCTTGGTG 187
603 a1SerValLysG1YAsnArgSerLysG1UAsnHisAlaAspLeuG1YLe 619
|||||
188 TCAGTGTCAAAAGTAAACGGTCAAAAGAACACGACGAGATTTGGGAATC 237
620 PheValLysSer1Le1LeaSnG1YG1YAla1a1aSerLysAspG1YArgLe 636
|||||
238 TTTTGCAAGTCATTTATTAATGAGAGACACACATCTTAAGATGGAAGCT 287
636 uArgValAsnAspG1nLeu1Lea1aValAsnG1YU1nSerLeuLeuG1YL 653
|||||
288 TCGGGTGAATGATCAACATGATAGCAGTAAATGGAATCCCTGTTGGGCA 337
653 ySThrAsnG1nAspAlaMetG1uThrLeuArgArgSerMetSerThrL 669
|||||
338 AGACAAACCAAGATGCCATGGAACCTTAAGAAGGTCATGCTACCTGAA 387
670 G1YAsnLysArgG1YMet1LeG1nLeu1LeValAlaArgArg1LeSerL 686
|||||
388 GGCATATAACGAGGATGATCCACCTTTATGTTGCAAGAGACAAATAGCAA 437
686 sCysAsnG1nLeuLysSerProG1YSerProProG1YProG1nLeuProI 703

Note: this is a NIH-MGC Library."

BASE COUNT 213 a 167 c 206 g 127 t

ORIGIN

alignment_scores:

Quality: 153.00 Length: 227.

Ratio: 0.674 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x BG170486 ..

Align seg 1/1 to: BG170486 from: 1 to: 713

```

1034 lysllleqluysrthrglylslllysllleqlngluserpherhrrsergl 1050
|||||
2 AAGATTGAGAAAACGGGTAAAAATTAATACAGAAATCCTTACATCAGA 51
1050 uglugluaryliearMetLysglnglulgluarylieqlnAlalyst 1067
|||||
52 AGAGAGAGGATACGAATGAGAGCAGAGAGAGAGATTCACGCCAAA 101
1067 hrarvgluPhearvgluarGlnAlarGluarqaspyrrAlaGluile 1083
|||||
102 CTGAGAAATTAGGAGACGACAGACTGAGAGCGCTGACTATGCTGAATT 151
1084 GlnAspPheHisArgThrPheGlyCysAspAspDluuueetyrGlyl 1100
|||||
152 CAAGATTTCATCGACATTTGGCTGTGATGATGATTAATGATGGGGG 201
1100 yAlSerSerTyrgluGlySerMetAlaLeuAsnAlarProGlnserp 1117
|||||
202 AGTTTCTTATGAGAGTTCCATGGCTCTCAACGCTAGACCTCAGAGCC 251
1117 roArvgluGlyHisMetMetAspAlaLeutyrrAlaGlnValysLysPro 1133
|||||
252 CACGAGAAAGGCATATGATGATGATGC.TTGATGCCCCAAGTCAMAGACCG 300
1134 ArgAsnSerLysProSerProValAspSerAsnArgSerThrProSerAs 1150
|||||
301 CGGAAATCCAAACCTCACCCTGTAAGACATACAGATCACTCTACGAA 350
1150 nHlAspArgIleGlnArGleuArGlnGluPheGlnGlnAlaLysGlnA 1167
|||||
351 TCATGATCGGATACAGCGCTCGAGGCAAGATTTCAGCAAGCAAGCAAG 400
1167 spgluAspValGluAspArgArgArgThrTyrsrPhegluglnProTrp 1183
|||||
401 ATGAAAGATGAGAAAGATCTCGCGGACCTATAGTTTGAAGCAACCTGG 450
1184 ProAsnAlarProAlaThrGlnSerGlyArGHisSerValSerValG 1200
|||||
451 CCGAGCGCAAGCGCGGAGCAGAGCGGCGGACACTCGGTCTCGTGGGA 500
1200 uValGlnMetGlnArGlnArGlnGlnGluArGluSerSerGlnGlnA 1217
|||||
501 GGTGCAGATCAGCGCAGCGCAGAGAGCGGAGAGCTCCACAGAGG 550
1217 laglnArGlnTyrsrSerSerLeuProArGlnSerArgLysAsnAlaSer 1233
|||||
551 CCGAGCGCCAGTACAGTCTCTGCTCGGCAAAACAGGAAAAATGCGAGC 600
1234 SerValSerGlnAspSerTyrgluGlnAsnTyrsrProGlygluglyph 1250
|||||
601 TCGGTCTCCAGAGACTCTTGGGAGCAGAACTACTCCCTCGGGGAAGCTT 650
1250 eGlnSerAlaLysGluAsnProArGlyTyrsr 1260
|||||
651 CCAGAGTGCAGAAAGAAACCCAGGTACTCC 681
seq_name: gb_est2:BG745723

```

seq_documentation_block:

LOCUS BG745723 1040 bp mRNA linear EST 15-MAY-2001

DEFINITION 602723832T1 NIH-MGC_113 Homo sapiens cDNA clone IMAGE:4850526 3',

RNA sequence.

ACCESSION BG745723 GI:14056376

VERSION BG745723.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1040)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: Dr. Mark Watson

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Plate: LICM1691 row: 0 column: 07

High quality sequence start: 15

High quality sequence stop: 850.

FEATURES

source

1..1040

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4850526"

/clone_1ib="NIH-MGC_113"

/lab_host="DH10B (phage-resistant)"

/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 231 a 265 c 220 g 324 t

ORIGIN

alignment_scores:

Quality: 152.00 Length: 152

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x BG745723/rev ..

Align seg 1/1 to reverse of: BG745723 from: 1 to: 1040

```

873 serProSerArgAspValaGlyProSerLeuGlyLeuLysLysSerSerSe 889
|||||
577 TCTCCACACAGATGCTGGCTCTTCCCTGGGTCTGAAAGAACTCAAGCTC 528
1184 ProAsnAlarProAlaThrGlnSerGlyArGHisSerValSerValG 1200
|||||
527 ATTGGAGAGCTCTGCAGACCGCAGTTGCCAGAGTGACTTTGAATGGGATA 478
1200 uValGlnMetGlnArGlnArGlnGlnGluArGluSerSerGlnGlnA 1217
|||||
889 rLeuGluSerLeuGlnThrAlaValaLalagluValrThrLeuAsnGlyAspI 906
|||||
527 ATTGGAGAGCTCTGCAGACCGCAGTTGCCAGAGTGACTTTGAATGGGATA 478
1200 uValGlnMetGlnArGlnArGlnGlnGluArGluSerSerGlnGlnA 1217
|||||
906 leProPheHisArgProArGProArGtlelleArGlyArGlyLysAsn 922
|||||
477 TTCTTTTCATCTGTCACAGCGCGGAGATTAACAGAGCGAGGATGCAAT 428
1234 SerValSerGlnAspSerTyrgluGlnAsnTyrsrProGlygluglyph 1250
|||||
923 GluSerPheArgAlaAlaIleAspLysSerTyrrAspLysProAlaValAs 939
|||||
427 GAGACCTTCAGAGCTGCCATGACAAATCTTATGCTTAACCCGCGGTAGA 378
1250 eGlnSerAlaLysGluAsnProArGlyTyrsr 1260
|||||
939 pAspAspAspLuglyMetGluThrLeuGluGluAspThrGluGlnSers 956

```

```

|||||
377 TGAATGATGAAGGAGCATGGACCTTGGAAAGACACAGAGAAAGTT 328
956 earrSerGlyAArgGluSerValSerThrAlaSerAspGlnProSerHis 972
327 CAAGATCAGAGGAGAGACTGCTTATCCACACCCAGTGATCCCTCCAC 278
973 SerLeuGlnArgGlnMetAsnGlnYasnGlnLulysGlyAspLysThrAs 989
277 TCTCTGGAGAGCAAAATGAATGAAACCAAGAAAGGTATAGACTGA 228
989 pargLysLysAspLysThrGlyLysGlnLulysLysLysAspArgLysG 1006
227 TAGAAAAAGGATAAACTGGAAAAAGAAAGAAAGATAGATAGAG 178
1006 ludyAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAsp 1022
177 AGAAGCATTAATAATGAACCAAGAGGAGATGCTGAAGGGCTTGGAGAC 128
1023 MetPhe 1024
|||||
127 ATGTTC 122

```

seq_name: gb_est2:BG260547

seq_documentation_block:

LOCUS BG260547 966 bp mRNA linear EST 13-FEB-2001
 DEFINITION 60237096F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4480366 5',
 mRNA sequence.

ACCESSION BG260547

VERSION BG260547.1 GI:12770363

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 966)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

COMMENT Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

Clone sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LHAM10313 Row: O Column: 23

High quality sequence stop: 723.

Location/Qualifiers

FEATURES

source

1..966

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4480366"

/clone_id="NIH_MGC_93"

/tissue_type="transitional cell papilloma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: bladder; Vector: pCMV-SPORT6; Site: 1; NotI;

Site: 2; SalI: Cloned unidirectionally; oligo-dt primed.

Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

BASE COUNT

ORIGIN 282 a 209 c 271 g 204 t

alignment_scores:

Quality: 149.00

Ratio: 1.000

Percent Similarity: 100.000

Length: 149

Gaps: 0

Percent Identity: 100.000

alignment_block:
 US-09-757-781-2 x BG260547 ..

Align seg 1/1 to: BG260547 from: 1 to: 966

```

592 ProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValLysGlyAs 608
|||||
1 CCACCTTAATGATTCAGGATCTGCAGGCTTGGTGTGAGTGCACAAAGTAA 50
608 narGSerLysGluAsnHisAlaAspLeuGlyIlePheValLysSerIleI 625
|||||
51 CCGGTCAAAAAGAGAACCCAGCAGATTTGGCAATCTTTGCAAGTCCATTA 100
625 leasnGlyGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAspGln 641
|||||
101 TTATATGAGAGAGCAGCATTTAAAGATGGAAGGCTTGGGTGATATGTCAA 150
642 LeuIleAlaValAsnGlyLysLeuLeuGlyLysThrAsnGlnAspAl 658
|||||
151 CTGATGACGATTAATGAGAAATCCCTGTTGGCAAGACAAACCAAGATGC 200
658 aMetGluThrLeuArgArgSerMetSerThrGluGlyAsnLysArgGlyM 675
|||||
201 CARGAAACCTTAAGAAAGGCTATGCTTACTGAAGGCATTAACGAGAA 250
675 etIleGlnLeuIleValAlaArgArgIleSerLysCysasnGluLeuLys 691
|||||
251 TGATTCAGCTTATGTGTGCAAGAGAAATAGCAAGTGCATATGCTGAAG 300
692 SerProGlySerProProGlyProGluLeuProIleGluThrAlaLeuAs 708
|||||
301 TCACCTGGAGGCCCTTGACCTGAGCTGCCCATTTGAACACCGTTGGA 350
708 pAspArgGluArgArgIleSerHisSerLeuTySerGlyIleGluGlyL 725
|||||
351 TGAATGAGACGAGAGAAATTTCCATTCCTCTACAGTGGGATTTGAGGGGC 400
725 euAspGluSerProSerArgAsnAlaAlaLeuSerArgIleMetGly 740
|||||
401 TGAATGAAATCGGCCAGCAAAATGCTGCCCTCAGTAGTAGTAATGGGT 447

```

seq_name: gb_est2:BG170054

seq_documentation_block:

LOCUS BG170054 987 bp mRNA linear EST 06-FEB-2001
 DEFINITION 602321514F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4424954 5',
 mRNA sequence.

ACCESSION BG170054

VERSION BG170054.1 GI:12676757

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 987)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

COMMENT Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LHAM10169 Row: K Column: 03

High quality sequence start: 7

High quality sequence stop: 633.

Location/Qualifiers

FEATURES

source

1..987

/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:4424954"
/clone_lib="NIH_MGC-89"
/tissue_type="hypermephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: kidney; Vector: pCMV-SPORT6; site.1: NotI;
site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      297 a      234 c      235 g      221 t
ORIGIN
alignment_scores:
    quality: 147.00      length: 147
    ratio: 1.000      gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x BGI70054 ..
Align seg 1/1 to: BGI70054 from: 1 to: 987
150 GlyleuSerThrSerValSerAspSerAsnPheserSergluGluProse 166
|||||
121 GGCCCTCCACTCTGTCAGTAGATGTAATTTCTCTGAAGAGCCCTTC 170
|||||
166 FArlYsaAnProthArGTpSerThrThrlAglyPheleuYsglnA 183
|||||
171 AAGGAAATCCACACGCTGGTCAACACAGCTGGCTTCTCAAGCAGA 220
|||||
183 snThrAlaGlySerProLySThrCysAspArgLySLysAspGluAsnTyr 199
|||||
221 ACACGTGGCGGAGTCTCTAAACCTGCGACAGAGAAAGATGAAACATAC 270
|||||
200 ArgSerLeuProArgAspPthSerAsnTrpSerAsnGlnPheGlnArgAS 216
|||||
271 AGAAGCCCTCCGCGGAGTACTAGTAACGTCTAACCAATTTTCAGAGAGA 320
|||||
216 PAsnAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpL 233
|||||
321 CAATGCTGCTGCTCTCTGAGTGCACAGTACCCTAATGTTGGCAAGTGGC 370
|||||
233 euGluYsgLIngluGlnAspGluAspGlyThrGluGluAspAsnSerArg 249
|||||
371 TGGAGAAACAGAACAGAGATGAGATGGGACAGAAAGAGATTAACAGTCTG 420
|||||
250 ValGluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPh 266
|||||
421 GTTGAACTGTTGGACATGACACGCGGTTTGGAGCATATATACCAACTT 470
|||||
266 eSerLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysP 283
|||||
471 TTCTCTGATGATATGTTAAAGCTGCTGTAAGATCCCAACGATGGAGGCG 520
|||||
283 rOlEuGlyIleHisValValProPheSerAlaArgGlyGly 296
|||||
521 CTCTGGGAATCATGTAGTGCCTTCACTGCTCGAGGCGGA 561
seq_name: gb_estl:A1990048
seq_documentation_block:
LOCUS      A1990048      442 bp      mRNA      linear      EST 08-MAR-2000
DEFINITION      ws19601.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2497656 3',
                mRNA sequence.
ACCESSION      A1990048
VERSION        A1990048.1      GI:5836929
KEYWORDS
SOURCE        human.
ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE      1 (bases 1 to 442)
AUTHORS
TITLE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL
COMMENT      Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LINL at:
            www.bio.lnll.gov/bdrrp/image/image.html
            Insert length: 1105      Std Error: 0.00
            Seq primer: -400P from Glibco
            High quality sequence stop: 430.
            Location/Qualifiers
                1..442
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2497656"
                /clone_lib="NCI_CGAP_GC6"
                /tissue_type="pooled germ cell tumors"
                /lab_host="DH10B"
                /Note="Vector: p773D-Pac (Pharmacia) with a modified
                polylinker. Site.1: Not I; Site.2: Eco RI; Plasmid DNA
                from the normalized library NCI-CGAP_GC4 was prepared, and
                ss circles were made in vitro. Following HAP purification,
                this DNA was used as tracer in a subtractive hybridization
                reaction. The driver was PCR-amplified cDNAs from a pool
                of 5,000 clones made from the same library (clones IDs
                1257096-1258631, 1469064-1470983, and 1475592-1476743).
                Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      88 a      120 c      87 g      147 t
ORIGIN
alignment_scores:
    quality: 145.00      length: 145
    ratio: 1.000      gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x A1990048/rev ..
Align seg 1/1 to reverse of: A1990048 from: 1 to: 442
845 ValLysThrArgLysSerLysSerMetAspLeuGlyIleAlaAspGluTh 861
|||||
442 GTTAAACACGAAATCAAAAAGCATGGATTAGGATATGCTGACGAGAGC 393
|||||
861 rLysLeuAsnThrValAspAspGlnLysAlaGlySerProSerArgAspY 878
|||||
392 TAAACTCAATACAGTGAATGACCAAGAACAGAGTCTCCACAGAGATG 343
|||||
878 aLgLyProSerLeuGlyLeuLysLysSerSerLeuGlnSerLeuGln 894
|||||
342 TGGGTCTCTCCCTGGGCTCTGAAGAAAGTCAAGCTCTGTGAGAGTCTGCG 293
|||||
895 ThrAlaValAlaGluValThrLeuAsnGlnGlyAspIleProPheHisArgPr 911
|||||
292 ACCGCAAGTGGCGAGGTGACTTGAATGGGATATTCCTTTCATCGTCC 243
|||||
911 cArgProArgIleIleArgGlyArgGlyCysAsnGluSerPheArgAlaA 928
|||||
242 ACGGCCGCGGATATACAGAGCGAGGAGATGCAATAGAGACTTCACAGAGCTG 193
|||||
928 IaIleAspLysSerTyrAspLysProAlaValAspAspAspGlyGly 944
|||||
192 CCATGACAAATCTTATGATTAACCCGCGGTAGATGATGATGATGAAGGC 143

```

```
945 MetGluThrLeuGluGluAspThrGluGluSerArgSerGlyArgI 961
|||||
142 ATGGAGACCTTGGAGAGACACAGAGAAAGTTCAAGATCAGCGAGACA 93
|||||
961 userValserThrAlaSerAspGlnProSerHisSerLeuGluArgGlnM 978
|||||
92 GTCTGATCCACAGCCAGTGCATCCCTCCCTCTCTGGAGAGACAAA 43
|||||
978 etAsnGlyAsnGlnGluLysGlyAspLysThrAsp 989
|||||
42 TGAATGGAACCAAGAGAAAGTGATTAAGACTGAT 8
```

OM of: US-09-757-781-2 to: Issued_Patents_NA:* out_format : pfs
Date: Jul 24, 2002 6:12 PM
About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=Frame+2.p2n.model -DEV=x1h
-Q=/cgn2_1/USPRO.spool/US09757781/runat_22072002.154317.18613/app_query.fasta.1.1434
-DB=Issued_Patents_NA -ORFMT=fastap -SUFFIX=olig.rni
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPTOL=0.000
-LOOEXT=0.000 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
-YGAPEXT=60.000 -DELDP=6.000 -DELEXT=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCCALIGN=200
-THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORF=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=20000000
-USER=US09757781 -ECGN1_1.68 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-757-781-2
Query length: 1356
Database: Issued_Patents_NA*
Database sequences: 383533
Database length: 122816752
Search time (sec): 65.160000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:
Strd Orig ZScore EScore Len Documentation
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-726-531-77 + 19.00 296.09 8.3e-09 1042
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-735-041A-3 + 10.00 137.71 5.53 1812
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-190-476B-3 + 10.00 137.71 5.53 1812
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-190-889A-3 - 10.00 137.71 5.53 1812
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-190-938B-3 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-735-041A-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-190-476B-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-190-889A-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-190-889A-7 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-190-938B-5 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-190-938B-7 - 10.00 136.60 6.37 2103
/cgn2_6/ptodata/2/1na/5A_COMB.seq:US-07-743-518-23 + 8.00 115.26 98.46 369
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-927-219-42 + 8.00 110.70 259.86 1021
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-818-112-20 + 8.00 107.70 259.86 1021
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-818-111-20 - 8.00 107.70 259.86 1021
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-056-556-20 - 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-611-757-33 + 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-611-757-33 - 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/1na/PCUTS_COMB.seq:PCUTS-05950-33 + 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/1na/PCUTS_COMB.seq:PCUTS-05950-33 - 8.00 107.40 269.80 1062
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-928-213B-58 - 8.00 107.08 281.17 1109
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-445-640-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-170-558-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-447-314-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-445-461-5 + 8.00 106.72 294.45 1164
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-004-171-1 - 8.00 106.28 311.56 1235
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-255-368-7 + 8.00 105.59 325.49 1293
/cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-415-657-14 + 8.00 105.59 325.49 1293
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-936-165A-113 - 8.00 105.10 362.60 1448
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-665-037-1 + 8.00 104.16 408.79 1642
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-666-067-1 + 8.00 104.16 408.79 1642
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-732-870-1 + 8.00 104.16 408.79 1642
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-466-954A-1 + 8.00 100.87 623.63 2557
/cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-286-305A-2 + 8.00 100.35 666.58 2742
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-441-104A-2 + 8.00 100.35 666.58 2742
/cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-440-816A-2 + 8.00 100.35 666.58 2742

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-276-531-77
seq_documentation_block:
Sequence 77, Application US/09276531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESSES:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,531
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/079,677
FILING DATE: March 27, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lynn E. Murry, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 1042 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BLADTUT04
CLONE: 1555118
US-09-276-531-77

alignment_scores:
Quality: 19.00 Length: 19
Percent Similarity: 100.000 Percent Identity: 100.000 Gaps: 0
alignment_block:
US-09-757-781-2 x US-09-276-531-77
US-09-757-781-2 ..
Align seg 1/1 to: US-09-276-531-77 from: 1 to: 1042

882 LeuglyleuLysSerSerLeugluSerLeugluInPhraIaVal1 898
|||||
662 CTGGCTTGAAAAAGTCCAGCTCTTGAGAGCTGCACAGCTGCGTGGC 711
898 aglVal1 900
|||||
712 CGAGGTC 718

seq_name: /cgn2_6/ptodata/2/lna/5b_COMB.seq:US-08-735-041A-3

seq_documentation block:

Sequence 3, Application US/08735041A
Patent No. 5914251

GENERAL INFORMATION:

APPLICANT: Farrell, Catherine L.
APPLICANT: Martin, Francis H.
APPLICANT: Yabkowitz, Rachel
TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH
TITLE OF INVENTION: FACTOR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:

ADDRESSEE: Amgen Inc.

STREET: 1840 De Havilland Drive

CITY: Thousand Oaks

STATE: California

COUNTRY: USA

ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/735,041A

FILING DATE: 22-OCT-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mazza, Richard J.

REGISTRATION NUMBER: 27,657

REFERENCE/DOCKET NUMBER: A-414

TELECOMMUNICATION INFORMATION:

TELEPHONE: 805.447.4112

TELEFAX: 805.499.6751

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1812 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-735-041A-3

alignment_scores:

Quality: 10.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x US-08-735-041A-3/rev ..

Align seg 1/1 to reverse of: US-08-735-041A-3 from: 1 to: 1812

420 ArgleuproHisSerAlaHisProSerGly 429

|||||

1681 CGGCTGCCACATTCGACATCCCTCGGG 1652

seq_name: /cgn2_6/ptodata/2/lna/6a_COMB.seq:US-09-190-476B-3

seq_documentation block:

Sequence 3, Application US/09190476B

Patent No. 6025204

GENERAL INFORMATION:

APPLICANT: Farrell, Catherine L.

APPLICANT: Martin, Francis H.

APPLICANT: Yabkowitz, Rachel

TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH

TITLE OF INVENTION: FACTOR

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amgen Inc.

STREET: 1840 De Havilland Drive

CITY: Thousand Oaks

STATE: California

COUNTRY: USA

ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/190,476B

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/735,041

FILING DATE: 22-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mazza, Richard J.

REGISTRATION NUMBER: 27,657

REFERENCE/DOCKET NUMBER: A-414

TELECOMMUNICATION INFORMATION:

TELEPHONE: 805.447.4112

TELEFAX: 805.499.6751

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1812 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-190-476B-3

alignment_scores:

Quality: 10.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x US-09-190-476B-3/rev ..

Align seg 1/1 to reverse of: US-09-190-476B-3 from: 1 to: 1812

420 ArgleuproHisSerAlaHisProSerGly 429

|||||

1681 CGGCTGCCACATTCGACATCCCTCGGG 1652

seq_name: /cgn2_6/ptodata/2/lna/6a_COMB.seq:US-09-190-889A-3

seq_documentation block:

Sequence 3, Application US/09190889A

Patent No. 6075008

GENERAL INFORMATION:

APPLICANT: Farrell, Catherine L.

APPLICANT: Martin, Francis H.

APPLICANT: Yabkowitz, Rachel

TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH

TITLE OF INVENTION: FACTOR

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amgen Inc.

STREET: 1840 De Havilland Drive

CITY: Thousand Oaks

STATE: California

```
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/190,889A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/735,041
FILING DATE: 22-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mazza, Richard J.
REGISTRATION NUMBER: 27,657
REFERENCE/DOCKET NUMBER: A-414
TELEPHONE: 805.447.4112
TELEFAX: 805.499.6751
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-190-889A-3

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-889A-3/rev ..
Align seg 1/1 to reverse of: US-09-190-889A-3 from: 1 to: 1812

420 ArgleuproHisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATTCGACATCCCTCCGGG 1652

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-190-938B-3

seq_documentation_block:
; Sequence 3, Application US/09190938B
; Patent No. 6197939
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; Martin, Francis H.
; Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,938B
; FILING DATE: 12-NOV-1998
; CLASSIFICATION: <unknown>
```

```
ATTORNEY/AGENT INFORMATION:
NAME: Mazza, Richard J.
REGISTRATION NUMBER: 27,657
REFERENCE/DOCKET NUMBER: A-414
TELEPHONE: 805.447.4112
TELEFAX: 805.499.6751
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-190-938B-3

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-938B-3/rev ..
Align seg 1/1 to reverse of: US-09-190-938B-3 from: 1 to: 1812

420 ArgleuproHisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATTCGACATCCCTCCGGG 1652

seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-735-041A-5

seq_documentation_block:
; Sequence 5, Application US/08735041A
; Patent No. 5914251
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; Martin, Francis H.
; Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,041A
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

MOLECULE TYPE: CDNA
US-08-735-041A-5

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-08-735-041A-5/rev ..

Align seg 1/1 to reverse of: US-08-735-041A-5 from: 1 to: 2103

420 ArgleuproHisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATCTCTGCACATCCCTCCGG 1652

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq:US-08-735-041A-7

seq_documentation_block:
; Sequence 7, Application US/08735041A
; Patent No. 5914251
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; APPLICANT: Martin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735, 041A
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805,447,4112
; TELEFAX: 805,499,6751
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-735-041A-7

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x US-08-735-041A-7/rev ..
Align seg 1/1 to reverse of: US-08-735-041A-7 from: 1 to: 2103

420 ArgleuproHisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATCTCTGCACATCCCTCCGG 1652

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-190-476B-5

seq_documentation_block:
; Sequence 5, Application US/09190476B
; Patent No. 6025204
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; APPLICANT: Martin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,476B
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/735,041
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805,447,4112
; TELEFAX: 805,499,6751
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-09-190-476B-5

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-476B-5/rev ..

Align seg 1/1 to reverse of: US-09-190-476B-5 from: 1 to: 2103

420 ArgleuproHisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATCTCTGCACATCCCTCCGG 1652

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-190-476B-7

seq_documentation_block:
; Sequence 7, Application US/09190476B
; Patent No. 6025204
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.

```

; APPLICANT: Martin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,476B
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/735,041
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-190-476B-7

alignment_scores:
      Quality: 10.00      Length: 10
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-476B-7/rev ..
Align seg 1/1 to reverse of: US-09-190-476B-7 from: 1 to: 2103

420 ArgLeupProHisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATCTGCACATCCCTCCGGG 1652

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-09-190-889A-5

seq_documentation_block:
; Sequence 5, Application US/09190889A
; Patent No. 6075008
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; APPLICANT: Martin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
```

```

; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,889A
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/735,041
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mazza, Richard J.
; REGISTRATION NUMBER: 27,657
; REFERENCE/DOCKET NUMBER: A-414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 805.447.4112
; TELEFAX: 805.499.6751
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-190-889A-5

alignment_scores:
      Quality: 10.00      Length: 10
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-889A-5/rev ..
Align seg 1/1 to reverse of: US-09-190-889A-5 from: 1 to: 2103

420 ArgLeupProHisSerAlaHisProSerGly 429
|||||
1681 CGGCTGCCACATCTGCACATCCCTCCGGG 1652

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:US-09-190-889A-7

seq_documentation_block:
; Sequence 7, Application US/09190889A
; Patent No. 6075008
; GENERAL INFORMATION:
; APPLICANT: Farrell, Catherine L.
; APPLICANT: Martin, Francis H.
; APPLICANT: Yabkowitz, Rachel
; TITLE OF INVENTION: PLACENTAL-DERIVED PROSTRATE GROWTH
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 De Havilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/190,889A
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
```

APPLICATION NUMBER: US 08/735,041
FILING DATE: 22-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mazza, Richard J.
REGISTRATION NUMBER: 27,657
REFERENCE/DOCKET NUMBER: A-414
TELECOMMUNICATION INFORMATION:
TELEPHONE: 805 447,4112
TELEFAX: 805 499,6751
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2103 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-190-889A-7

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-889A-7/rev ..

Align seg 1/1 to reverse of: US-09-190-889A-7 from: 1 to: 2103

420 ArgleuproHISeralahisProsergly 429
|||||
1681 CGGCTGCCACATCTCTGCACATCCCTCCGG 1652

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-190-938B-5

seq_documentation_block:
Sequence 5, Application US/09190938B
Patent No. 6197939
GENERAL INFORMATION:
APPLICANT: Farrell, Catherine L.
Martin, Francis H.
Yabkowitz, Rachel
TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH
FACTOR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/190,938B
FILING DATE: 12-NO. 6197939-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mazza, Richard J.
REGISTRATION NUMBER: 27,657
REFERENCE/DOCKET NUMBER: A-414
TELECOMMUNICATION INFORMATION:
TELEPHONE: 805 447,4112
TELEFAX: 805 499,6751
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2103 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-190-938B-5

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-938B-5/rev ..

Align seg 1/1 to reverse of: US-09-190-938B-5 from: 1 to: 2103

420 ArgleuproHISeralahisProsergly 429
|||||
1681 CGGCTGCCACATCTCTGCACATCCCTCCGG 1652

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-190-938B-7

seq_documentation_block:
Sequence 7, Application US/09190938B
Patent No. 6197939
GENERAL INFORMATION:
APPLICANT: Farrell, Catherine L.
Martin, Francis H.
Yabkowitz, Rachel
TITLE OF INVENTION: PLACENTAL-DERIVED PROSTATE GROWTH
FACTOR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/190,938B
FILING DATE: 12-NO. 6197939-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mazza, Richard J.
REGISTRATION NUMBER: 27,657
REFERENCE/DOCKET NUMBER: A-414
TELECOMMUNICATION INFORMATION:
TELEPHONE: 805 447,4112
TELEFAX: 805 499,6751
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2103 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-190-938B-7

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x US-09-190-938B-7/rev ..

Align seg 1/1 to reverse of: US-09-190-938B-7 from: 1 to: 2103

420 ArgLeuProHisSerAlaHisProSergly 429
|||||
1681 CGGCTGCCACATTTCTGCACATCCCTCCGGG 1652

seq_name: /cgn2_6/ptodata/2/lna/5A.COMB.seq:US-07-743-518-23

seq_documentation_block:

; Sequence 23, Application US/07743518
; Patent No. 5397696
; GENERAL INFORMATION:
; APPLICANT: YAMAGIHARA, RICHARD
; APPLICANT: NERURKAR, VIVEK R.
; APPLICANT: JENKINS, CAROL
; APPLICANT: MILLER, MARK
; APPLICANT: GARRUTO, RALPH M.
; TITLE OF INVENTION: PAPUA NEW GUINEA HUMAN T-LYMPHOTROPIC
; TITLE OF INVENTION: VIRUS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,518
; FILING DATE: 19910812
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: WTS/5683/84699/SAP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 369 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..343
; US-07-743-518-23

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x US-07-743-518-23 ..

Align seg 1/1 to: US-07-743-518-23 from: 1 to: 369

766 LeuProValLeuProHisIleu 773
|||||
7 CTTCCAGTTTCGCCCCGCATCTT 30

seq_name: /cgn2_6/ptodata/2/lna/6B.COMB.seq:US-08-927-219-42

seq_documentation_block:

; Sequence 42, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisha
; APPLICANT: Kalsaki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,219
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,679
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION NUMBER: US 60/028,056
; FILING DATE: 02-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,719
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 662 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 84..188
; US-08-927-219-42

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-757-781-2 x US-08-927-219-42 ..

Align seg 1/1 to: US-08-927-219-42 from: 1 to: 662

767 ProValLeuProHisIleuser 774
|||||
46 CCGTGTCCCTCCCTCACCTCTCT 69

Thu Jul 25 08:38:22 2002

us-09-757-781-2.olg.rni


```

873 rProSerArgAspValGlyProSerLeuGlyLeuLysLysSerSerL 890
|||||
2349 TCCAGACAGAGATGTGGGTCTCTCCCTGGGTGAAAGACTCAAGCTCGT 2398
890 euGluSerLeuGlnThrAlaValAlaValThrLeuAsnGlyAspIle 906
2399 TGGAGAGCTGCAGACCGCAAGTTCGAGGTGCTTGAATGGGATATT 2448
907 ProPheHisArgProArgProArgIleIleArgGlyArgGlyCysAsnG 923
|||||
2449 CTTTCATCTGTCACGCGCCGCGATATATCAGAGCGAGGGATGCAATGA 2498
923 userPheArgAlaIleIleAspLysSerTyrAspLysProAlaValAsp 940
|||||
2499 GAGCTTCAGAGCTGCCATCGACAAATCTTATGATTAACCCCGGTAGATG 2548
940 sPaAspGluGlyMetGluThrLeuGluGluAspThrGluGluSerSer 956
|||||
2549 ATGATGATGAGGCATGGAGACCTTGGAAAGACACAGAAAGAGTTCA 2598
957 ArgSerGlyArgGluSerValSerThrAlaSerAspGlnProSerHis 973
|||||
2599 AGATCAGGAGAGAGACTGTATCCACAGCCAGATGATCAGCTCCACTC 2648
973 rLeuGluArgGluMetAsnGlyAsnGlnGlyAspLysThrAsp 990
|||||
2649 TCTGGAGACAAATGAAATGAAACCAAGAGAAAGGTGATAGACTGATA 2698
990 rGlyLysAspLysThrGlyLysGluLysLysLysAspArgAspLysGlu 1006
|||||
2699 GAAAAAGCATAAACTGGAAGAAAGAAAGAAAGATAGAGATTAAGAGAG 2748
1007 LysAspLysMetLysAlaLysLysGlyMetLeuGlyLeuGlyAspMe 1023
|||||
2749 AAGGATTAATGAAAGCCAAAGAGGAATGCTGAAGGCTTGGAGACAT 2798
1023 tPheArgPheGlyLysHisArgLysAspAspLysIleGluLysThrGly 1040
|||||
2799 GTTCAGGTTTGGCAACATCGAAAAAGATGACAAAGATTGAGAAACGGGTA 2848
1040 yEiLysIleGlnGluSerPheThrSerGluGluGluLysIleArgMet 1056
2849 AATTAATAATACAGGAATCTTTACATCAGAAAGAGAGGATTCAGAAATG 2898
1057 LysGlnGluGlnGluArgIleGlnAlaLysThrArgGluPheArgLys 1073
2899 AAGCAGAGACGAGAGAGATTCAGCCAAACTCGAGAAATTTAGGGAACG 2948
1073 gGlnAlaArgGluArgAspTyrAlaGluIleGlnAspPheHisArgThr 1090
2949 ACAAGCTCAGAGCGCTGACTATGTAATTCAGAAATTTTCATCGGACAT 2998
1090 heGlyCysAspAspGluLeuMetTyrGlyGlyValSerSerTyrGluGly 1106
|||||
2999 TTGGCTGTGATGATGATTAATGATGGGAGCTTCTCTATGAAGGT 3048
1107 SerMetAlaLeuAsnAlaArgProGlnSerProArgGluGlyHisMet 1123
|||||
3049 TCCATGGCTTCACAGCTAGACCTCAAGCCACGAGAAAGGCGCATATGAT 3098
1123 tAspAlaLeuTyrAlaGlnValLysLysProArgAsnSerLysProSer 1140
3099 GGATGCTTTGATGCCCAAGTCAAGAGCCGCGAATTCCAAAACCTTCAC 3148
1140 roValAspSerAsnArgSerThrProSerAsnHisAspArgIleGlnArg 1156
3149 CTGTAGACAGTACAGATCACTCTACCAATCATGATCGGATACACCT 3198
1157 LeuArgGlnGluPheGlnAlaLysGlnAspGluAspValGluAspArg 1173
3199 CTGAGGCAAGAAATTTACGCAAGCAAGCAAGATGAATATGAAGATCG 3248
1173 gArgArgThrTyrSerPheGluGlnProTyrProAsnAlaArgProAla 1190

```

```

3249 |||||
3249 TCGGGGACCTTATAGTTTGGACACCTGGCCGAACGCGCGCGCGCA 3298
1190 hrgInSerGlyArgHisSerValSerValGluValGlnMetGlnArgGln 1206
|||||
3299 CGCAGACGCGGCGACACTCGGTGCTCCGTGAGAGTGCAGATCGACGCGAC 3348
1207 ArgGlnGluGluArgGluSerSerGlnAlaGlnArgGlnTyrSerSe 1223
3349 CGGAGAGAGAGGCGCAGAGCTCCACAGCAGGCCACGCGCATGATACACT 3398
1223 rLeuProArgGlnSerArgLysAsnAlaSerSerValSerGlnAspSer 1240
3399 TCTGCTCCGCAAGACAGAAATTCACAGCTCGCTCTCCACAGACTCT 1240
1240 rPglGluAsnTyrSerProGlyGluGlyPheGlnSerAlaLysGluAsn 1256
3449 GCGACACACACTACTCCCTGGGGAAGGCTTCCAGAGTCCCAAGAGAAC 1257
1257 ProArgTyrSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyLys 1273
3499 CCCAGTACTCTCAGCTACAGGCTCCAGAAAGGCTACTGCGAGAGCA 1273
1273 sGlyPheAsnAlaArgValMetLeuGluThrGlnGluLeuLeuArgGln 1290
3549 TGGCTTCACGCGCAGGCTCATCTGGAACACAGAGAGCTCTTCCAGG 1290
1290 LngInArgArgLysGluGlnGlnMetLysLysGlnProProSerGluGly 1306
3599 AACAGAGCGGAGAGACGACGATGAAAGACAGCCCTTCTCCAGGGG 1307
1307 ProSerAsnTyrAspSerTyrLysLysValGlnAspProSerTyrAla 1323
3649 CCCACACTATGACTGTATAGAAAGTCCAGAGCCAGTACGCCCC 1323
1323 oProLysGlyProPheArgGlnAspValProProSerProSerGlnVal 1340
3699 TCCCAAGGGCCCTTCCGCAAGATGTCCCCCTCCCTTCACGGTTG 1340
1340 lArgGluAsnArgLeuGlnThrProGluLysGlyArgProPheTyrSer 1356
3749 CGAGGCTGAACAGACTTCCAGACTCTCGAAGAAAGGAGGCCCTTATAT 1356
seq_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH18587
seq_documentation_block:
ID AAH18587 standard; CDNA: 2718 BP.
AC
AAH18587;
XX
26-JUN-2001 (first entry)
DT
XX
DE Human cDNA sequence SEQ ID NO:18774.
XX
Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX
EP1074617-A2.
PD
XX
07-FEB-2001.
XX
28-JUL-2000; 2000EP-0116126.
PF
XX
29-JUL-1999; 99JP-0248036.
PR
27-AUG-1999; 99JP-0300253.
PR
11-JAN-2000; 2000JP-0118776.
PR
02-MAY-2000; 2000JP-0183767.
PR
09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

```

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8: SEQ ID 18774; 2537p + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other:
SQ
alignment_scores:
Quality: 420.00 Length: 508
Ratio: 0.830 Gaps: 1
Percent Similarity: 99.606 Percent Identity: 99.606
alignment_block:
US-09-757-781-2 x AAH18587 ..
Align seg 1/1 to: AAH18587 from: 1 to: 2718
234 GIUUGSGINGUGINASPGLIASPGLYTHRGUGUASPASNSERARGVA 250
|||||
327 GAGAAACAGACAGACGATGAGATGGACAGAAAGGATTAACAGTCGTCT 376
250 IGLUPROVALGLYHISALAAPRTHRGILEUGLIHISILLEPROASNPHE 267
|||||
377 TGAACCTGTTGACATGCTGCACAGGGTTTGGAGCATATACCAACTTTT 426
267 ERLEUASPAPMETVALYLSLEUVALGLUVALPROASNPGLYGLYPRO 283
|||||
427 CTCTGGATGATATGTAAAGCTCTAGAAATGCCCAACGATGAGAGGCT 476
284 LEUGLILEHISVALVALPROPHESERIALARGLYGLYARGYTHLEUGI 300
|||||
477 CTGGGAATCCATGTAGTGCCTTTCACTGCTCAGGCGGACAGAACCTGGG 526
300 YLEUGLEUVALYLSARGLEUGILYLSGLYGLYLSALAGLIHISGLUASN 317
|||||
527 GTTATTTAGTAAACGATTTGAGAAAGTGTAAGCTGAACATGAATAATC 576
317 EUPHEIARGGLUASNAPCYSTILEVALARGYLEASNAPGLYASPLEUARG 333
|||||
577 TTTTTCGTGAGAAATGATTCATGTGCAGGATTAATGATGGACCTTGA 626
334 ASNARGARGPHEGLUGLINALAGLIHISERPHARGGLINALEMARGTH 350
|||||

627 AATAGAACATTTGAACAACACACATATGTTCCCAAGCCATCCGTAC 676
350 PROILEIETRPHEHISVALVALPROALAALAASNLYSGIUNGITRG 367
|||||
677 ACCCATATTTGGTTCATGTGTGCTCGACCAATATAAGAGCATNG 726
367 IUGLILEUSERGINSERGLYUASNASNTYTYRSERARPHESER 383
|||||
727 AACACATATCCCAAGTAGAAGAACAAATTAATATCAAGCCGTTTAC 776
384 PROASPSERGINTRYLLEASPAPNARGSERVALANSERIALAGLYLEUHI 400
|||||
777 CCTGACACCCAGTATATTTGACAAACAGAGTGTGAACAGTCAGGCGCTTCA 826
400 STHVALGLNARGALAPROARGLEUASNHISPROFUGIUNGILNLEASPS 417
827 CACGCTGCAGAGACACCCCGACTACACCCGCTGACGAGATNAGCT 876
417 ERHISSEARGLEUPROHISSERIALAHISPROSEGLYLYSPROPSER 433
|||||
877 CTCACCTCAAGACTACCTCATAGCGCACACCCCTCGGAAACACCATCC 926
434 ALAPROALASERALAPROGLINASNVALPHESERTHTRVALSERSEGI 450
927 GCTCCAGCCTCGGACCTCAGATGATATTTACTACGACTGTAAAGCATGG 976
450 YTYRASNTHRLYSILLEGILYLSARGLEUASNILEGLINLEUYSLYSG 467
977 TTATATNACCAAAAAATATGCGCAAGAGCTTATATTCAGCTTAAGAAAG 1026
467 IYTHGLUGILEUGLYPHESERIIERTHSERARASPVATHRILEGILY 483
1027 GTACAGAAAGGTTTGGATTCACATCCTCCAGAGATGATTAACAATAGCT 1076
484 GLYSERIALAPROILETYRVALYASNIILEUPROARGLYVALALAL 500
1077 GCTCAGCTCCAGTATGTGAAAAACATTCCTCCCGGGGGCGCCCTT 1126
500 EGLINAPGLYARGLEUYSALAGLYASPARGLEUILEGLUVALAENGILYV 517
1127 TCAGGATGGCCGACTTAAGCGAGACAGACTTATAGAGGTAAATGCGAG 1176
517 ALASPLEUVALGLYSSERGINGLUGLUVALVALSERLEUENARGSER 533
|||||
1177 TAGATTTAGTGGCAAAATCCCAAGAGGAAGTGTTCGCTGTGGAAGAC 1226
534 THRLYSMETGLUGLYTHRVALSERLEUVALPHEARGGLUGLUSP 550
1227 ACCAAGATGGAAAGAACTGTGAGCCCTTCTGTTCCTCCAGGAAGACGC 1276
550 APHEHISPROARGIULEUASNALAGLUPROSERGINMETGINLEPROL 567
1277 CTTCACCCCAAGGGAATGAAATGAGAGCAAGCCAGATGCAATTCGA 1326
567 YSGIUTHRLYSALAGLUSPGLIUSPILLEVALLLEUTHPROASNPGLYTHR 583
1327 AAGAAACGAAAGCAAGAGTGAAGTATGTCTTACACCTGATGGACACC 1376
584 ARGGLUPHELEUTHRPHLEGLUVALPROLEUASNAPSERGLYSERIALAG 600
1377 AAGGAATTTCTGACATTTGAACTCCCACTTAAGTCAATTCAGAGATGGCAG 1425
600 IYLEUGLIVALSERVALYLSGLYASNARGSERLYSGIUNHISALAASP 616
1426 GCGTTGGGTGTCAGTCAAGGTAAACGGTCAAAAAGAGAACACCGAGAT 1475
617 LEUGLIYIIEPHEVALYSSERTLEILEANGILYGLYALALAISERLYSAS 633
1476 TTGGAAATCTTTGTCAAGTCCATTATTAATGAGAGAGCGCATCTAAAGA 1525
633 PGLYARGLEUARGVALASNAPGLINLEUILEVALVALSANGILYUGISERL 650
1526 TGGAGGCTTCGGGTGAATGATCACTGACATGACAGTAAATGGAGATCC 1575

```

650 euleuglylvsrhrasnglnaspalametglutthleuargargsermet 666
    |||||||
1576 TGTGGGCAAGACAAACCAAGATGCATGAAACCTTAAGAAAGGCTCTATG 1625
    |||||||
667 SerThgluylvsrhrasnglymetllleuilevalalaaargar 683
    |||||||
1626 TCACAGCAAGCAATACAGCAAGCAATGACGCTTATGTTGCAAGGAG 1675
    |||||||
683 gllserlyscyasngluileuylserproglyserproglyserprogly 700
    |||||||
1676 AATACCAAGTCAATGACGCTGACGCTGACGCTGACGCTGACGCTG 1725
    |||||||
700 luleuprolleugluthralaleuaspsargluarargylleserhis 716
    |||||||
1726 AGCTGCCATTGAAACAGCGTTGATGATGAGAACAGAAATTTCCCAT 1775
    |||||||
717 Serleuylrserglyllegluylleuaspsgluierproglyserglnal 733
    |||||||
1776 TCCTCTACAGTGGGATTTGAGGGGCTTGATGATGATGATGATGATG 1825
    |||||||
733 AlaleuSerArglleMetgly 740
    |||||||
1826 TGCCCTCAGTAGATGAATGCGT 1847

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH78369
seq_documentation_block:
ID   AAH78369 standard; DNA; 2718 BP.
AC   AAH78369;
XX
DT   26-NOV-2001 (first entry)
DE   Nucleotide sequence of a human protein kinase/protein phosphatase.
XX
KW   Human; protein kinase; protein phosphatase; signal transduction; ss.
XX
OS   Homo sapiens.
XX
FH   Key
FT   CDS
FT   Location/Qualifiers
FT     33..2630
FT     /*tag= a
FT     /product= "protein kinase/protein phosphatase"
XX
PN   M0200109316-A1.
XX
PD   08-FEB-2001.
XX
PE   28-JUL-2000; 2000MO-IP05061.
XX
PF   29-JUL-1999; 99UP-0248036.
XX
PR   18-OCT-1999; 99US-0159590.
XX
PR   11-JAN-2000; 2000JP-0118776.
XX
PR   17-FEB-2000; 2000US-0183322.
XX
PR   02-MAY-2000; 2000UP-0183767.
XX
PR   09-JUN-2000; 2000UP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
PI   Senoo C, Nezu J;
XX
XX   MPI: 2001-570286/64.
XX   P-PSDB; AAG67610.
XX
XX   New genes encoding proteins with protein kinase/protein phosphatase
XX   activity, useful in the diagnosis and treatment of diseases -
XX   Claim 1, Page 67-76; 233pp; Japanese.
XX
XX   The present sequence encodes a human protein kinase/protein phosphatase.

```

```

CC   It is expected that the protein kinase/protein phosphatase gene
CC   participates in signal transduction in cells. The protein kinase/protein
CC   phosphatase polypeptides and polynucleotides are useful for developing
CC   diagnostics and treatment agents for human and animal diseases. The
CC   protein kinase/protein phosphatase polypeptides are useful as target
CC   molecules in designing novel drugs. The protein kinase/protein
CC   phosphatase polynucleotides are useful as a source of probes and
CC   primers, which may be used to isolate homologous sequences.
XX
XX   Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other.

alignment_scores:
  Quality: 420.00      Length: 508
  Ratio: 0.830
  Percent Similarity: 99.606      Percent Identity: 99.606

alignment_block:
US-09-757-781-2 x AAH78369 ..

Align seq 1/1 to: AAH78369 from: 1 to: 2718

234  Gluylsglnluclnaspcluaspglythrngluaspasnerargva 250
    |||||||
327  GAGAACCAAGAACAGAGATGAGATGGACAGAAAGATTAACAGTCGTCT 376
    |||||||
250  lgluprovalglyhisalaspthrnglyleugluhisleproasphes 267
    |||||||
377  TGAACCTGTTGGACATGCTGACACGGGTTTGGACGATATACCAACTTTT 426
    |||||||
267  erleuaspsmetvalylsleuvalgluvalproaspsglglypro 283
    |||||||
427  CTCGTGATATATGTTAAAGCTCGTAGAAATGCCCAAGATGAGGGCCT 476
    |||||||
284  leuglyllehisvalvalpropheseralarglylarythrleugl 300
    |||||||
477  CTGGGAATCCANTGTAGTCTCTTCACTGCTCTGAGCGCGAGAACCTCGGG 526
    |||||||
300  yleuuevalylsarargleuglylsglyllysalaagluhisgluasnl 317
    |||||||
527  GTTATTACTTAACGATTTGGAGAAAGGTGGTAAAGCTGAACATGAAATTC 576
    |||||||
317  euphearggluasnaspcysllevalarglleasnasprglyasprleuarg 333
    |||||||
577  TTTTCGTGAGAAATGATTGCTGATGATGATGATGATGATGATGATGATG 626
    |||||||
334  Asnargrarphneglucinalaglnhismetphearglnalameargth 350
    |||||||
627  AATGAGAGATTTGACACAGCAACATATGTTCCCAAGCCATCGTAC 676
    |||||||
350  rprolleiletrpnehisvalvalproalalaasnlsgluinltyrg 367
    |||||||
677  ACCCATCATTTGGTTCCATGTGGTTCTCTACACAAATTAAGGCGATGTG 726
    |||||||
367  lueuileuSerGlnSerGluylsasnasntyrtyrSerArgpheser 383
    |||||||
727  AACAACTATCCCAAGTGAAGAAACAATTACTATTCAAGCGCTTTAGC 776
    |||||||
384  ProAspSerGlnTyrIleAspsnarGserValasnserralaglyleuNI 400
    |||||||
777  CCGAGACGCCAGTATATTTGACACAGCGAGTGTGAACAGTGGAGGGCTTCA 826
    |||||||
400  sthrvalglinargalaproh-gleuasnhisproprogluclnleasps 417
    |||||||
827  CACGGTGCAGAGACGCCCGACCTGAACACCCCGTACAGAGATAGACT 876
    |||||||
417  ernhisSerArgleuprohisSerIahisproserglylvspropoSer 433
    |||||||
877  CTCACCTCAAGACTACCTATAGCCACACCCCTCGGAAAAACCATCATCC 926
    |||||||
434  AlaproAlaseralaproglnasnlvalpheserthrThrValSerSerGcl 450
    |||||||
927  GCTCAGCCTCGGACCCAGAAATGTATTAGTACGACTGTAAAGCAGTGG 976

```

```

450 YTYRAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysG 467
|||||
977 TTYTAAACACCAAAAAATAGCCAAAGAGCTTAAATATCCAGCTTAAGAAAG 1026
|||||
467 LyrThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 483
|||||
1027 GTTACGAAAGGTTTGGGATCTCAGCATCTCCACAGAGATGTAAACAATAGGT 1076
|||||
484 GlySerAlaProIleTyrValLysAsnIleLeuProArgIleAlaAlaI 500
|||||
1077 GGCTCAGACTCCAAATCTATGTGAAAAACATTCTCCCGGGGGGGGGCCAT 1126
|||||
500 eGlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGly 517
|||||
1127 TCAGGATGGCCGACTTAAGCAGAGACAGACTTATAGAGCTAAATGAG 1176
|||||
517 AlaPheValuAlaGlyLysSerGlnGluValLysIleLeuAsnArgSer 533
|||||
1177 TAGATTATAGTGGCAAAATCCCAAGAGAAAGTTGTTGCTGTGTGAGAAAGC 1226
|||||
534 ThrLysMetGluGlyThrValSerLeuLeuValPheArgGlnIleAspAl 550
|||||
1227 ACCAAGATGGAGAGAACTGTGAGCCTCTGTGCTTGGCCAGAAAGCC 1276
|||||
550 aPheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProI 567
|||||
1277 CTTCACCCCAAGGAGAACTGAATCAGAGCCAAAGCCAGATCGAATCCAA 1326
|||||
567 ySGIuThrLysAlaGluAspGluAspIleValIleuThrProAspGlyThr 583
|||||
1327 AAGAAGACAAAGACAGAGATGAGATGTGTTCTTACACCGAATGGAGCAC 1376
|||||
584 ArgGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaG 600
|||||
1377 AGGGAATTTCTGACATTGGAAGTCCCACTTAGT.GATTCAAGATCTGAG 1425
|||||
600 LysGluGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAsp 616
|||||
1426 GCCTTGGTGTGCAGTGCAAAGGTAAACCGTCAAAAGAGAAACACGACAGAT 1475
|||||
617 LeuGlyIlePheValLysSerIleIleAsnGlyGlyValAlaIleSerLys 633
|||||
1476 TTGGGAATCTTTGTCAAGTCCATTTATATGAGAGAGACAGATCTAAAGA 1525
|||||
633 pGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerL 650
|||||
1526 TGGAAAGCCTTCGGGTGATGATCAACTGATGACAGTAATGAGAGATCCC 1575
|||||
650 euleuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMet 666
|||||
1576 TGTGTGGGCAAGACAAACCAAGATCCATGGAACCTTAAGAAAGTCATATG 1625
|||||
667 SerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaIleArg 683
|||||
1626 TCTTCTGAAGGCATTAACGAGGATGATCCAGCTTATGTGCAAGAGAG 1675
|||||
683 gIleSerLysCysAsnGluLeuLysSerProGlySerProProGlyProG 700
|||||
1676 AATTAAGCAAGTGAATGAGTGAAGTCACTGGGAGCCCCCTGGAGCTG 1725
|||||
700 LuleuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHis 716
|||||
1726 AGCGGCCCATTTGAAACAGCGTTGGATGATAGACAGAAAGAAATTTCCCAT 1775
|||||
717 SerLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAl 733
|||||
1776 TCCCTCTACAGTGGGATTTGAGGGGCTTGATGATATGCCACAGCAAAATGC 1825
|||||
733 AlaLeuSerArgIleMetGly 740
|||||
1826 TGCCCTCAGTGAATGAATGGGT 1847

```

```

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AAH72793
seq_documentation_block:
ID_AAH72793 standard; cDNA; 1800 BP.
XX
AC_AAH72793;
XX
DT_19-SEP-2001 (first entry)
XX
DE_Human cervical cancer marker nucleic acid 4067.
XX
KM_Cervical cancer: cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS_Homo sapiens.
XX
PN_W0200142467-A2.
XX
PD_14-JUN-2001.
XX
PF_08-DEC-2000; 2000MO-US33312.
XX
PR_08-DEC-1999; 99US-0169681.
PR_21-DEC-1999; 99US-0171350.
PR_14-MAR-2000; 2000US-0189315.
PR_12-MAY-2000; 2000US-0203791.
PR_09-JUN-2000; 2000US-0210600.
PR_21-JUL-2000; 2000US-0220114.
XX
PA_(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI_Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR_WPI; 2001-375006/39.
XX
PT_New isolated nucleic acid for diagnosing and treating cervical cancer
PT_and for assessing and detecting compounds for treating the cancer -
XX
PS_Claim 1: Page 855-856; 1051pp; English.
XX
CC_The invention relates to novel genes (AAH68727-AAH73383) associated with
CC_cervical cancer with cytostatic activity. The nucleic acids and encoded
CC_polypeptides are useful: to assess if a patient is afflicted with
CC_cervical cancer or has a pre-malignant condition; to monitor the
CC_progression of cervical cancer or a premalignant condition in a patient;
CC_to select and/or assess the efficacy of a compound or therapy for
CC_inhibiting cervical cancer in a patient. The nucleic acids may also be
CC_useful for gene therapy.
XX
SO_Sequence 1800 BP; 578 A; 359 C; 466 G; 394 T; 3 other;

alignment_scores:
Quality: 310.00 Length: 553
Ratio: 0.564 Gaps: 1
Percent Similarity: 99.458 Percent Identity: 99.458

alignment_block:
US-09-757-781-2 x AAH72793 ..
Align seg 1/1 to: AAH72793 from: 1 to: 1800

472 GlyPheSerIleThrSerArgAspValThrIleGlyGlySerAlaProI 488
|||||
4 GGATTCACACATCACTTCCAGAGATGTAAACAATAGGTGCTCAGCTCCGAT 53
|||||
488 eTyrValLysAsnIleLeuProArgGlyValAlaIleGlnAspGlyArgL 505
|||||
54 CTATGTGAAACATTTCTCCCGGGGGGGCCCATTCAGATGGCCGAGC 103
|||||
505 eulysAlaGlyAspArgLeuIleGluValAsnGlyValAspLeuValGly 521
|||||
104 TTAAGGCAAGAGACAGACTTATAGAGTAAATGAGTGAATTTATATGGGC 153

```

522 LysSerGlnGluValSerLeuArgSerThrLysMetGlu 538
154 AAATCCCAAGAGAGTTGTCCTGTTGACAGACCAAGATGGAAG 203
538 YThrValSerLeuValPheArgGlnIuAspAlaPheHisProArg 555
204 AACTGAGAGCTTGTGCTTTCGACAGAGAGAGCTTCCACCCAGAG 253
555 IuLeuAsnAlaGluProSerGlnMetGlnIleProLysGluThrLysAla 571
254 AACTGATGCAGACCAAGCCAGATGCAGATTCCAAAGAAACCAAGCA 303
572 GluAspGluAspIleValLeuThrProAspGlyThrArgGluPheLeu 588
304 GAAGATGAGAGATATTGTTCTTACACCTGATGACACAGGGAATTCGAC 353
588 rPheGluValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSer 605
354 ATTGAGTCCCACTTATGATTCAGATCTGCGGCTTGCTGACAGT 603
605 allYsGlyAsnArgSerLysGluAsnHisAlaAspLeuGlyIlePheVal 621
404 TCMAAGGTACCCGGTCAAAAGAGAACCCAGATTTGGGAATCTTGTG 453
622 LysSerIleIleAsnGlyAlaAlaSerLysAspGlyArgLeuArg 638
454 AAGTCCATTATTATGAGAGAGAGCATCTMAAGATGGAAGGCTGGGT 503
638 LAsnAspGlnLeuIleAlaValAsnGlyLysLeuLeuGlyLysThr 655
504 GAATGATCAACTGATACGATGATGAAATGGAATCCCTGTGGGCAAGCA 553
655 snGlnAspAlaMetGluThrLeuArgArgSerMetSerThrGlnLys 671
554 ACCAATATGCCATGAAACCTTAAGAGTCTATCTCTACTGAAAGGCAT 603
672 LysArgGlyMetIleGlnLeuIleValAlaArgArgLysSerLys 688
604 AACGAGGATGATCCAGCTTATCTTCCAAAGAAATACCAAGATGCA 653
688 nGluLeuLysSerProGlySerProGlyProGluLeuProIleGlu 705
654 TGAAGTGAAGTCACTGGAGAGCCCGGACCTGAGCTGCCCTTGA 703
705 hrAlaLeuAspAspArgGluArgArgLysSerHisSerLeuYrSer 721
704 CAGCTTGATGATGAGAACGAAATTTCCATCTCTACAGAGGG 753
722 IleGluGlyLeuAspGluSerProSerArgAsnAlaIleSerArg 738
754 ATTCAGGGGCTTGATGATGATGCCAGCAAAATGCTCCCTCAGTAGAT 803
738 eMetGlyLysSerGlyLysThrGlnLeuSerProThrValAsnMetPro 755
804 AATG.....GTTAAATACCAAGCTGCTCCCTACAGTGAATATGCCCC 844
755 LAspAspThrValIleIleGluAspAspArgLeuProValLeuPro 771
845 AAGATGACACTGATATGAGATGACAGGTGCAATGCTCTCCCA 894
772 HisLeuSerAspGlnSerSerSerSerHisAspAspValGlyPhe 788
895 CACTCTCTACAGTCTCTCTCCAGCTCCCATGATGATGGGGTTGT 944
788 IThrAlaAspAlaGlyThrTrpAlaLysAlaIleSerAspSerAla 805
945 GAGGAGAGATGCTGGTACTTGGCCAAAGCTGCAATCATGATTCAGCG 994
805 spcYsSerLeuSerProAspValAspProValLeuAlaPheGlnArg 821
995 ACTGCTCTTGAAGTCCAGATGTTGATCCAGTTCGCTTTCACAGGAA 1044
822 GlyPheGlyArgGlnSerMetSerGluLysArgThrLysGlnPheSer 838

1045 GGAATTGGACGTACAGATATGCAAGAAAAGCAAAAGCAATTTTCA 1094
838 PALASerGlnLeuAspPheValLysThrArgLysSerLysSerMetAsp 855
1095 TCCAGTCAATTTGGATTTCGTTAAACACGAAATCAAAAAGCATGATT 1144
855 euGlyIleAlaAspGluThrLysLeuAsnThrValAspAspGlnLysAla 871
1145 TAGCTATAGCTGACAGACATAAATCAATACAGTGCAGACGAAAGCA 1194
872 GlySerProSerArgAspValGlyProSerLeuGlyLeuLysSerSe 888
1195 GTTCTCTCCAGACAGATGTGGGTCTTCCCTGGCTGTAAGATGCAAG 1244
888 rSerLeuGluSerLeuGlnThrAlaValAlaGluValThrLeuAsnGly 905
1245 CTCGTTGAGAGTCTCAGACCGCAGTCCAGAGTGCAGCTTGAAATGGG 1294
905 spLleProPheHisArgProArgProArgIleIleArgGlyArgGlyCys 921
1295 ATATTCTCTTCCATTCCTCCACGCGCGGATATCAGAGGAGGGATGC 1344
922 AsnGluSerPheArgAlaAlaIleAspLysSerYrAspLysProAla 938
1345 AATGAGAGCTTCAGAGCTCCATCGACAAATCTTATGATTAACCCGCGT 1394
938 LAspAspAspAspGluGlyMetGluThrLeuGluGluAspThrGluLys 955
1395 AGATGATGATGATGAAAGCATGAGACCTTGAAAGACACAGAGAA 1444
955 eSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnProSer 971
1445 GTTCAAGATCAGGAGAGAGTGTATCCACAGCCAGTATGACCTTCC 1494
972 HisSerLeuGluArgGlnMetAsnGlyAsnGlnIuLysGlyAspLys 988
1495 CACTCTGAGAGACAAATGATGAAACCAAGCAAGAGATGATTAAGAC 1544
988 rAspArgLysLysAspLysThrGlyLysGluLysLysLysAspArgAsp 1005
1545 TGATGAAAAAAGATTAACCTGAAAAAGAAAAGAAAGATAGATGA 1594
1005 YsGluLysAspLysMetLysAlaLysGlyMetLeuLysGlyLeuGly 1021
1595 AGGAGAAAGATTAATAAGAAAGCCAAAGAGGAATGCTGAAGGCTTGGGA 1644
1022 AspMetPhe 1024
1645 GACATGTTCC 1653
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAS27320
seq_documentation_block:
ID AAS27320 standard; CDNA: 823 BP.
AC AAS27320;
DI 07-NOV-2001 (first entry)
DE CDNA encoding novel signal transduction pathway protein, Seq ID 355.
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.

OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0186874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

xx Rosen CA, Barash SC, Ruben SM;
 PI
 xx WPI: 2001-465460/50.
 xx
 DR P-PSDB: AA017403.
 xx
 xx Novel polypeptides useful for diagnosing, treating, preventing and/or
 PT prognosing disorders related to the proteins, including cancers, immune
 PT disorders and neuronal disorders -
 xx
 xx Claim 1: SEQ ID No 355; 880pp; English.

xx The invention relates to novel isolated polypeptides (I), and
cc polynucleotides (II). (I), (II) and the antibody to (I) are useful for
cc diagnosing, preventing and treating diseases including immune system
cc disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
cc disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
cc transplant rejections and graft versus host disease, infectious diseases
cc (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
cc other blood-related disorders (sickle cell anaemia), myeloproliferative
cc disorders, primary haematopoietic disorders, hyperproliferative
cc disorders (e.g. Gaucher's disease and cancer), neurodegenerative
cc disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
cc abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
cc disorders (e.g. glomerulonephritis), cardiovascular disorders
cc disorders (e.g. glomerulonephritis), respiratory disorders, endocrine disorders (e.g.
cc (e.g. arrhythmia), epithelial cell proliferation, gastrointestinal
cc wound healing, epithelial cell proliferation, endocrine disorders (e.g.
cc Addison's disease), reproductive system disorders (cystitis),
cc disorder (inflammatory disorders), liver disorders (cirrhosis),
cc as stimulators of B-cell responsiveness to pathogens, activators of
cc T-cells, to induce higher affinity antibodies, and as a means to induce
cc tumour proliferation in pathological e.g. acquired immune deficiency
cc syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
cc pathway protein coding sequences and PCR primers of the invention.

```
alignment_scores:      Length: 195
                      Quality:
                        Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
```

alignment_block: US-09-757-781-2 x AAS27320 ..

Align seg 1/1 to: AAS27320 from: 1 to: 823

4223	CysTleValIrtgIleasnspGlyspIleuhtArgasnArqAqPheglugI	339
195	TGCAATTGTCCAGGAATTATGATGTGGCGGACCTTGCAAAATACAGACTTTGAAACA	244
339	nalaglnHisMetPheArgGlnIleMetArqThrProIleIetPhePh	356
245	AGCCAAACATATGTGTTGCCAAACCATGGGTACACCCATCATTTGGTTCC	294
356	IsValValProAlaAlaAlaAsnlyGluGlnItyrTrbIubIubIeuSerGlnSer	372
295	ANGTGGTTCCTGCAGCAATATTAAGACGATGATACACTATTCGCCAAAT	344
373	GluItyrAsnAsnItyrTyrSerSerArgPheSerProAspSerGlnItyrTl	389
345	GAGGAGAACAAATCTATTATTCAGCCCTTTTACGCCCTTGACGCCACGATATAT	394
389	eAspAsnAcqSerValIasnSerAlaGlyLeuIlnIsthrValGIAlArAlAr	406
395	TGCAACACAGAGTGTAAACAGTGCAGGGCTTCACACGGTCGACAGAGAC	444
406	roArqLeuAsnHisProPrcGluGlnIlnIleAspSerHisSerArgLeuPro	422
445	CCCGACTGAACCAACCCGCTGGAGCAGATAGACTGCATCTCAAGACTACT	494
423	HisSerIlnHisProSerGlyItyrProPrcSerAlarProAlaSerAlar	433
495	CATAGCCACACACCCCTCGGAGAAACACCATCTCCGCTCCAGCCTCGGACAC	544

```

439  oGlnAsnValPheSerThrValSerSergLYTYAsnThrIlySI 456
      |||||
545  TCAGAAATGATATTAGTACGACGTAGCAGAGGTGTATTAACCAAAAAA 594
      |||||
456  IeGLIyATgLeAsnIIeGLInLeIySgIyThrGluIyLeuGI 472
      |||||
595  YAGGCAAGAGGCTTAAATATCCACTTAAGAAAAGATACAGAGGTGGGA 644
      |||||
473  PheSerIIeThrSerATgAspValThrIeGIyGIySerAlaProIley 489
      |||||
645  TTCAGATCATCTCCAGACAGATGTAAACAATAGTGGCTCAGCTCCAAATCA 694
      |||||
489  rValIyAsnIIeLeuProATgGIyAlaAlaIeGLInSpGIyATgLeuL 506
      |||||
695  TGTGAAAAACATTCCTCCCGCGGGGGGGCGGCATTCAGATGGCCGACCTTA 744
      |||||
506  ySalAGIyAspATgLeuIIeGLIyAlAsnGIyVal 517
      |||||
745  AGCGACGACAGACACTTATAGAGGTAAATGGACATA 779
      |||||

```

```
seq_name: /SIBSI/gcgucac/...
seq_documentation_block:
ID   AAH33594 standard; cDNA; 823 BP
```

XX
AC
AAH33594;

XX
DT 03-SEP-2001 (first entry)

XX	antigen encoding	CDNA SEQ ID NO:650
XX		

DE Human colon cancer antigen; diagnosis; detection
XX

Human; colon cancer; colorectal carcinoma; ss

XX
OS Homo sapiens

XX
PN WO200122920-A2

XX
PD 05-APR-2001.

XX
PF 28-SEP-2000; 2000WO-US26524.

XX	99US-0157137
PR	99US-0163280
	29-SEP-1999;
	03-NOV-1999;

PR 03-NOV-1999, 2000
XX
XX
(HUMAN GENOME SCI INC.

PA (HOMER) HOMER
XX Tuben SM Barash SC, Birse CE, Rosen CA

PI RUBEN SM/ 2001-235357/24
XX
ST

DR WFL, 2001-01-01
DR P-PSDB; AAG74163.
DR
XX

Claim 1: Page 2675; 98039P; English.

PS cancer-associated proteins (PS) and cancer-associated nucleic acid molecules (N) and proteins (P), where
XX AAH37195 and AAG77514 to AAG77788 represent human colon
CC AAH372943 to AAH37195 and AAG77514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC associated with decreased expression by rectifying mutations or deletions
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing P.
CC Inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell.
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the

CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 823 BP; 249 A; 183 C; 201 G; 188 T; 2 other:

alignment_scores:
Quality: 195.00 Length: 195
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x AAF33594 ..

Align seg 1/1 to: AAF33594 from: 1 to: 823

```
323 CysIleValArgIleAsnAspGlyAspLeuArgAsnArgPheGluG1 339
|||||
195 TGCATTCGACGATTATGATGCGCGACCTTCGAATAGAAATGTTGAACA 244
339 nAlaGlnHisMetPheArgGlnAlaMetArgThrProIleIleThrPheH 356
|||||
245 AGCACAACATATGTTGCGCAAGCCATGCGTACACCCATCATTTGGTCC 294
356 IsValValProAlaAlaAsnLysGluGlnIleArgGlnSer 372
|||||
295 ATGGGTTCGCGCAAAATTAAGACAGATATGAACAATGCCAAGT 344
373 GluLysAsnAsnIleTyrSerSerArgPheSerProAspSerGlnIle 389
|||||
345 GAGAGAGACAACTACTATTCAGCCGTTTACCCCTGACACCAGTATAT 394
389 eAspAsnArgSerValAsnSerAlaGlyLeuHisThrValGlnArgAlaP 406
|||||
395 TGACACAAGAGGTGTGAACAGTGCAGGCTTCACACGGTGCAGAGACAC 444
406 roATgLeuAsnHisProProGluGlnIleAspSerHisSerArgLeuPro 422
|||||
445 CCCAGTGAACACCCCGCTGACAGATAGACTCTCACTCAAACTACCT 494
423 HisSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaP 439
|||||
495 CATAGCCGACACCCCTCGGAAACACACATCCCTCGACCTCGGACAC 544
439 oGlnAsnValPheSerThrThrValSerSerGlyTyrAsnThrIleLysI 456
|||||
545 TCAGAAATGTTATTTAGTACGACTGTAACAGTGGTTATTAACCCAAAAA 594
456 leGlyLysArgLeuAsnIleGlnLeuLysLysGlyThrGluGlyLeuGly 472
|||||
595 TAGCAGAAGAGCTTAATATCCAGCTTAAGAAAGTACGAAGGTTTGGA 644
473 PheSerIleThrSerArgAspValThrIleGlyGlySerAlaProIleTy 489
|||||
645 TTCAGCATCTCTCCAGAGATGAATAGTGGTCTCAGCTCCAACTTA 694
489 rValLysAsnIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuL 506
|||||
695 TGTGAAAAAACCTTCTCCCGCGGGGGCGGCATTCAGGATGGCCGACTTA 744
506 ysaLagIysArgLeuIleGluValAsnGlyVal 517
|||||
745 AGCAGAGACAGACTTATAGAGTAAATGAGTA 779
```

seq_name: /STDJ/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF67583

seq_documentation_block:
ID AAF67583 standard; cDNA; 425 BP.

XX
AC AAF67583;
XX

DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 3345.
XX

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX

OS Homo sapiens.
XX

PN MO200102568-A2.
XX

PD 11-JAN-2001.
XX

XX 30-JUN-2000; 2000MO-US18374.
XX

PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX

PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Ckenjakov R, Drmanac S, Dickson M, Labat I, Leschkowitz D;
PI Kita D, Garcia V, Jones LW, Strache-Grain B;
XX

DR WPI; 2001-091805/10.
XX

XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 1045; 1046pp; English.
XX

CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX

SQ Sequence 425 BP; 134 A; 86 C; 110 G; 95 T; 0 other;

alignment_scores:
Quality: 131.00 Length: 131
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-757-781-2 x AAF67583 ..

Align seg 1/1 to: AAF67583 from: 1 to: 425

```
586 PheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLeuG1 602
|||||
33 TTTCTGACATTTGAAGTCCACCTTAATGATTCAGATCTGCAGGCTTGG 82
602 yValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuGlyI 619
|||||
83 TGTCAAGTGTCAAAAGGTACCGGTCAAAAGACACACCAAGATTTGGGAA 132
619 lePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspGlyArg 635
|||||
```


XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG04716.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS
 PS Claim 1; SEQ ID NO 4707; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human.
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SO Sequence 1262 BP; 405 A; 246 C; 307 G; 304 T; 0 other;

alignment_scores:
 Quality: 92.00 Length: 120
 Ratio: 0.780 Gaps: 2
 Percent Similarity: 98.333 Percent Identity: 98.333

alignment_block:
 US-09-757-781-2 x AAS68903 ..

Align seg 1/1 to: AAS68903 from: 1 to: 1262

1026 PheGlyLysHisArgLysAspAspLysIleGluLysThrGlyLysIle 1042
 337 TTGGGCAAACTCGAAAAAGATGACAAATTCAGAAACGGGTAAATAA 386
 1042 sileGlnLysSerPheThrSerGluGluGluArgIleArgMetLysGln 1059
 387 AATACAGGAATCCTTACATCAGAAAGAGAGATAGCATGAAGCAGG 436
 1059 LugiGluArgIleGlnAlaLysThrArgGluPheArgGln. ArgGlnAl 1075
 437 AGCAGGAGAGATTCAGCCAAACCTCGAATTTAGGGAACCCAGCAGC 486
 1075 AARGGluArgSerPheArgIleGlnAlaPheHisArgThrPheGlyC 1092
 487 TCGAGAGGAGTACTATGCTGAATTCAGATTTTCATCGGACATTTGGCT 536
 1092 ysaAspAspGluLeuMetLysGlyLysValSerSerTyGluGlySerMet 1108
 537 GTGATGATGAGTAACTATGATGGGGAGATTCTTCTATGAAGGTTCCATG 586

1109 AlaleuAsnAlaArgProGlnSerProArgGluGlyHisMetMetAspAl 1125
 587 GCCTCAACGCTAGACCTCAGACCCAGCAAGGCGATATGATGATGTC 636
 1125 a.LeuTyraIleGlnValLysLysProArgAsnSerLysProSerProVal 1141
 637 CTTTGTAATGCCCAAGTCAGAGAGCGGGAATTCAGAACCTCAGCTGTA 686
 1142 AspSer 1143
 687 GACAGC 692

seq_name: /STDS1/gcgdata/hold_geneseg/genesegn-emb1/AA2001A.DAT:AAH71480

seq_documentation_block:
 ID AAH71480 standard; cDNA; 597 BP.

AC AAH71480;
 DT 19-SEP-2001 (first entry)
 DE Human cervical cancer marker nucleic acid 2754.
 KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
 OS Homo sapiens.
 PN WO200142467-A2.
 PD 14-JUN-2001.
 PF 08-DEC-2000; 2000WO-US33312.
 PR 08-DEC-1999; 99US-0169681.
 PR 21-DEC-1999; 99US-0171350.
 PR 14-MAR-2000; 2000US-0189315.
 PR 12-MAY-2000; 2000US-0203791.
 PR 09-JUN-2000; 2000US-0210600.
 PR 21-JUL-2000; 2000US-0220114.
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PI Schlegel R, Deeds J, Berger A, Zhao X;
 DR WPI; 2001-375006/39.
 PT New isolated nucleic acid for diagnosing and treating cervical cancer
 PT and for assessing and detecting compounds for treating the cancer -
 PS
 PS Claim 1; Page 566; 1051bp; English.

CC The invention relates to novel genes (AAH68727-AAH73383) associated with
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded
 CC polypeptides are useful: to assess if a patient is afflicted with
 CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a premalignant condition for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.

SO Sequence 597 BP; 180 A; 123 C; 156 G; 138 T; 0 other;

alignment_scores:
 Quality: 89.00 Length: 89
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-757-781-2 x AAH71480 ..

Align seg 1/1 to: AAH71480 from: 1 to: 597

Thu Jul 25 08:38:21 2002

us-09-757-781-2.oli.g.rng

468 ThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly1 484
1 ACGAAGGTTGGGATTCAGCATCTCCAGACATGTAACTAGATG 50
484 YSRAlaProIleTyrValLysAsnIleLeuProArgGlyAlaIleG 501
51 CTGAGCTCCATATGTGCAAAACATTCCTCCCGGGGGCCGCTTC 100
501 InAspGlyArgLeuLysAlaGlyAspArgLeuIleGlyValAsnGlyVal 517
101 AGGATGGCCCGACTTAAGGCGAGAGAGCTTATAGAGTAAATGAGCTA 150
518 AspleuValGlyLysSerGlnGluValValSerLeuLeuArgSerTh 534
151 GATTAGTGGGCAATCCCAAGAGAGAGTGTTCCTTGTGAGAACGAC 200
534 rLysMetGluTyrThrValSerLeuLeuValPheArgGlnGluAspAla 551
201 CAAAGTGAAGAGACTGTGAGCTTGTGCTTTCGCGCAGAGAGACGCT 250
551 heHisProArgGluLeu 556
251 TCCACCCCAAGGAGACTG 267
seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF66057
seq_documentation_block:
ID AAF66057 standard; cDNA: 400 BP.
AC AAF66057;
CC 09-APR-2001 (first entry)
DE Novel human polynucleotide, SEQ ID NO: 1813.
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
OS Homo sapiens.
PN WO200102568-A2.
PD 11-JAN-2001.
PF 30-JUN-2000; 2000WO-US18374.
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
PI Williams LR, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones LM, Strache-Crain B;
XX WPI; 2001-091805/10.
XX Library of polynucleotides for diagnosing a cancerous state of a
XX mammalian cell and detecting cancer, particularly of the colon or
XX prostate, comprises 3351 human polynucleotide sequences -
PS Claim 9; Page 803; 1046pp; English.
XX The present sequence is one of 3351 sequences in a library of human
XX polynucleotides. The library is used to detect differentially expressed
XX genes correlated with a cancerous state of a mammalian cell and can
XX detect colon, prostate, breast and lung cancer. The library can be used
XX to produce probes for detection of mRNA and to produce additional copies
XX of the polynucleotides. The probes can be used for chromosome mapping of
XX the polynucleotide and for detection of transcription levels. Ribozymes
XX or antisense oligonucleotides can be generated. The polynucleotides and

CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SO Sequence 400 BP; 118 A; 85 C; 99 G; 98 T; 0 other;
alignment_scores:
Quality: 84.00 Length: 84
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AAF66057 ..
Align seg 1/1 to: AAF66057 from: 1 to: 400
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeu1 817
149 AGGATTCAGCCGACTGCTTGTGAGTCCAGATGTGATCACTTCTTGC 198
817 apheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
199 TTTTCAACGAGAGAGATTGGAGCTGCAGCATGATGCGAAAAACGCCAAC 248
834 YSGInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
249 AGCAATTTTCAGATGCCAGCAATTCGATTCGTAAACAGCAAAATCA 298
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
299 AAAAGCATGATTTAGGTATAGCTGACGAGACTAAACCTCAATACAGTGA 348
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
349 TGACCAGAAAGCAGGTTCTCCACGACAGATGTGGTCTCCCTCGGCGTC 398
884 eu 884
399 TG 400
seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AA568901
seq_documentation_block:
ID AA568901 standard; cDNA: 582 BP.
AC AA568901;
CC 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #4705.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI

seq_documentation_block:

ID AAD23370 standard; cDNA: 540 BP.

AC AAD23370;

XX 26-FEB-2002 (first entry)

DE Human lung tumour-specific SCC1-568 cDNA.

XX Human: lung tumour protein; immunostimulant; cytostatic; gene therapy;
XX antisense-therapy; vaccine; immune response; lung cancer; SCC1-568; ss.
OS Homo sapiens.

PN WO200172295-A2.

XX 04-OCT-2001.

PF 28-MAR-2001; 2001WO-US09991.

PR 29-MAR-2000; 2000US-0538037.

PR 05-JUN-2000; 2000US-0588937.

PR 18-AUG-2000; 2000US-0640878.

PR 22-SEP-2000; 2000US-234517P.

PR 01-NOV-2000; 2000US-0704512.

PR 14-DEC-2000; 2000US-0738973.

XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;

XX Henderson RA, Filing SP, Algate PA, Elliot M, Mannion J, Kalos MD;

XX WPI; 2001-639201/73.

XX New human lung-specific polynucleotides and polypeptides for the

XX diagnosis and treatment of disease e.g. lung cancer

XX Example 2; Page 276; 378pp; English.

XX The invention relates to isolated lung tumour-specific proteins and

XX their corresponding cDNA molecules. Lung tumour-specific proteins and

XX their antigen-presenting cells are useful for stimulating and/or

XX the development of cancer. The invention also relates to a composition

XX useful for stimulating an immune response, and for treating cancer.

XX CC lung tumour specific oligonucleotide is useful in gene therapy and for

XX CC diagnosis, detection and treatment of lung cancer. The present sequence

XX is human lung tumour-specific cDNA.

XX Sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other;

XX SQ

XX alignment_scores:

XX Quality: 78.00

XX Ratio: 1.000

XX Percent Similarity: 100.000

XX Percent Identity: 100.000

XX Alignment block:

XX US-09-757-781-2 x AAD23370

329 AGTCCCTTCGACGCCATGATGATGGGTTTGACGGCAGATGCT 378

793 GYTHTRPALALyAlaAlaIleSerAspSerAlaAspCysSerLeuSe 809

379 GGTACTTGCGCCAGAGCTGCATCAGTGAATCAGCCGACTGCTTTGAG 428

809 rProAspValAspProValLeuAlaPheGlnArg 820

429 TCAGATGTTGATCCAGTCTTCTTTCACGA 462

seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH07109

seq_documentation_block:

ID AAH07109 standard; cDNA: 556 BP.

AC AAH07109;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:3944.

XX Human: primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118176.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX full-length cDNAs

XX Claim 1; SEQ ID 3944; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

Thu Jul 25 08:38:21 2002

```
XX      Sequence 556 BP; 154 A; 129 C; 144 G; 126 T; 3 other;
SQ
alignment_scores:
    Quality: 64.00      Length: 64
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-757-781-2 x AAH07109 ..
Align seg 1/1 to: AAH07109 from: 1 to: 556
126 GluValThrProSerValLeuArgAlaAsnMetProLeuHisValArgAr 142
|||||
3 GAGGTACACACTTCAGTCCTCGAGCAAAATGCGCTCTCATGTTGAGCG 52
142 gSerSerAspProAlaLeuIleGlyLeuSerThrSerValSerAspSera 159
|||||
53 CAGTAGTGACCCAGCTCTAATGGCTCTCCACTTGTGTCAGTATAGTA 102
159 snpSerSerGluGluProSerArgLysAsnProThrArgTrpSerThr 175
|||||
103 ATTTTCTCTGAAGAGCCTTCAAGAAAAATCCACACGCTGGGTCAACA 152
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLys 189
|||||
153 ACAGCTGGCTTCTCAAGACAGACACTGCTGGAGTCTTAA 194
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:33:34 ; Search time 30.49 Seconds

(without alignments)
7693.707 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356
Sequence: 1 MKYVCFGRTRVVPCCGDH.....SOVARLNRLQTPKGRPEYS 1356

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size: 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1113	82.1	1353	4	Q9BY58 homo sapien
2	781	57.6	1031	4	Q9BY57 homo sapien
3	556	41.0	988	4	Q96RM7 homo sapien
4	556	41.0	1273	4	Q96RM6 homo sapien
5	500	36.9	1266	4	Q9NYE6 homo sapien
6	413	30.5	667	4	Q9HC48 homo sapien
7	406	29.9	865	4	Q96K28 homo sapien
8	347	25.6	347	4	Q9NM14 homo sapien
9	88	6.5	1337	11	Q9Z340 ratu mus
10	77	5.7	1333	11	Q9Z340 ratu mus
11	19	1.4	624	4	Q96DK9 homo sapien
12	19	1.4	1143	4	Q96NM6 homo sapien
13	19	1.4	1205	4	Q96NM7 homo sapien
14	13	1.0	545	4	Q96N09 homo sapien
15	12	0.9	128	6	Q95L72 macaca fasc
16	11	0.8	1464	5	Q96782 drosophila

17	11	0.8	1464	5	Q9VX75	Q9VX75 drosophila
18	11	0.8	2460	11	Q64512	Q64512 mus musculu
19	10	0.7	79	4	Q96Q29	Q96Q29 homo sapien
20	10	0.7	213	11	Q9CSB4	Q9CSB4 mus musculu
21	10	0.7	1017	16	Q9ZMD1	Q9ZMD1 rhizobium m
22	8	0.6	74	6	Q9TSZ5	Q9TSZ5 canis famill
23	8	0.6	88	2	Q9ZGY1	Q9ZGY1 yersinia pe
24	8	0.6	107	3	Q9Y813	Q9Y813 schizosacch
25	8	0.6	128	10	Q23020	Q23020 arabidopsis
26	8	0.6	153	16	P96120	P96120 treponema p
27	8	0.6	219	16	Q9K618	Q9K618 bacillus ha
28	8	0.6	226	11	Q9DCR6	Q9DCR6 mus musculu
29	8	0.6	287	10	Q9ZK23	Q9ZK23 triticum sp
30	8	0.6	296	4	Q14909	Q14909 homo sapien
31	8	0.6	297	5	P90976	P90976 caenorhabd
32	8	0.6	315	11	Q91Y70	Q91Y70 rattus norv
33	8	0.6	316	5	Q9U245	Q9U245 caenorhabd
34	8	0.6	318	5	Q22229	Q22229 caenorhabd
35	8	0.6	323	2	Q93Q60	Q93Q60 rhizobium m
36	8	0.6	326	4	Q00272	Q00272 homo sapien
37	8	0.6	334	11	Q9D3C6	Q9D3C6 mus musculu
38	8	0.6	335	2	Q9F2T7	Q9F2T7 streptomyce
39	8	0.6	335	11	Q91V06	Q91V06 mus musculu
40	8	0.6	336	16	Q92SA7	Q92SA7 rhizobium m
41	8	0.6	337	4	Q00556	Q00556 homo sapien
42	8	0.6	337	11	Q9JHL1	Q9JHL1 mus musculu
43	8	0.6	337	11	Q920G2	Q920G2 rattus norv
44	8	0.6	373	11	Q9CVL4	Q9CVL4 mus musculu
45	8	0.6	450	4	Q15599	Q15599 homo sapien

ALIGNMENTS

RESULT 1
ID Q9BY58 PRELIMINARY; PRT; 1353 AA.
AC Q9BY58;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN LONG VARIANT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID-9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fang C., Xu Y.,
RT "Exon/Intron Structure and Splicing Variants of a Novel Human Polarity
RT gene, HASIP.";
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF196185; AAK27891.1; -.
DR HSSP: Q12923; 3PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PSS0106; PDZ; 3.
SQ SEQUENCE 1353 AA; 151149 MW; BC531577B9C31AE3 CRC64;

Query Match 82.1%; Score 1113; DB 4; Length 1353;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1353; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 MKYVCFGRTRVVPCCGDHMKVFSLIQQA VTRRAIAKDPYWTQVRLRHGGDGIID 60
DB 1 MKYVCFGRTRVVPCCGDHMKVFSLIQQA VTRRAIAKDPYWTQVRLRHGGDGIID 60
QY 61 LDDIILCDVADDDKRLVA VDEDDPHHGGDGTASSGTGSPETFGSELGNNVASAPQPY 120
DB 61 LDDIILCDVADDDKRLVA VDEDDPHHGGDGTASSGTGSPETFGSELGNNVASAPQPY 120

```

OY 121 ANSEIVTSPVLNANMPLHVRSSDPALIGLSTVSQSNFSESPSKNPTRMSTTAGFL 180
DB 121 ANSELEVTPSVLRANMPLHVRSSDPALIGLSTVSQSNFSESPSKNPTRMSTTAGFL 180
OY 181 KONTAGSPPTCKDKKDEYRSLPRDTSNWSNOFORNARSLASASHPMVKMLEKOEODE 240
DB 181 KONTAGSPPTCKDKKDEYRSLPRDTSNWSNOFORNARSLASASHPMVKMLEKOEODE 240
OY 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGLGIHVVPFSAARGRTLG 300
DB 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGLGIHVVPFSAARGRTLG 300
OY 301 LLYKLEKGGKAHEHENLFRENDCLVRINDGLNRRREFEQAHMFROAMRTPIIMFHVPA 360
DB 301 LLYKLEKGGKAHEHENLFRENDCLVRINDGLNRRREFEQAHMFROAMRTPIIMFHVPA 360
OY 361 ANKEQYEQLSQSEKNYSSRSPDSQYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHR 420
DB 361 ANKEQYEQLSQSEKNYSSRSPDSQYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHR 420
OY 421 LPHSAHPGKPPASAPASQNVSTVSSGYNTKKIGKRLNQLKKGTEGLGFSTSRDV 480
DB 421 LPHSAHPGKPPASAPASQNVSTVSSGYNTKKIGKRLNQLKKGTEGLGFSTSRDV 480
OY 481 TIGGSAPITYKNTIPRGAIIODGRKAGBRLIEVNGVDLYKSOEYVSLSTRKMGTV 540
DB 481 TIGGSAPITYKNTIPRGAIIODGRKAGBRLIEVNGVDLYKSOEYVSLSTRKMGTV 540
OY 541 SLTVFROEAPFRELNEPSPOMQIPKTKAEDDIYLPDGTREFLTFEVPINDSSAG 600
DB 541 SLTVFROEAPFRELNEPSPOMQIPKTKAEDDIYLPDGTREFLTFEVPINDSSAG 600
OY 601 LGVSVKGNRSKENHADLGIKFKSIINGGAOKDRLRVNOLIAVNGESLIGTNDAME 660
DB 601 LGVSVKGNRSKENHADLGIKFKSIINGGAOKDRLRVNOLIAVNGESLIGTNDAME 660
OY 661 TLRRSMSTEGKRGKIOIIVARRISKNEKSPGSPPELPLEIADLRERISHSLYS 720
DB 661 TLRRSMSTEGKRGKIOIIVARRISKNEKSPGSPPELPLEIADLRERISHSLYS 720
OY 721 GIEGLDSSPSNMAALSRIMESGKYQSPVNNMPODDYIIEDDRLPYLPHLSQSSSS 780
DB 721 GIEGLDSSPSNMAALSRIMESGKYQSPVNNMPODDYIIEDDRLPYLPHLSQSSSS 780
OY 781 SHDVGFTVADAGTMAKAALISDADCSLSPDYVLAFOREGFGQSMSEKRTKOFSPAS 840
DB 781 SHDVGFTVADAGTMAKAALISDADCSLSPDYVLAFOREGFGQSMSEKRTKOFSPAS 840
OY 841 QIDFVYTRKSKMDLGIADETKLNIVDDQKAGSPRDYGPRLGLKSSLSLQTAFAEV 900
DB 841 QIDFVYTRKSKMDLGIADETKLNIVDDQKAGSPRDYGPRLGLKSSLSLQTAFAEV 900
OY 901 TLNGDIPFRRPRRIIRGCGNESFRAIDISYDKPAVDDDEGEMTLEEDTESSSGR 960
DB 901 TLNGDIPFRRPRRIIRGCGNESFRAIDISYDKPAVDDDEGEMTLEEDTESSSGR 960
OY 961 EASYSTRASDOPSHLERONGNOEGKDTDRKKDKTGKEKKDKDRBEKDKAKKMLKGL 1020
DB 961 EASYSTRASDOPSHLERONGNOEGKDTDRKKDKTGKEKKDKDRBEKDKAKKMLKGL 1020
OY 1021 GDMFFGKRRKDDKIEKGIKIKIOESTSEEBERTRMQOEORIOAKTRERERQARERDY 1080
DB 1021 GDMFFGKRRKDDKIEKGIKIKIOESTSEEBERTRMQOEORIOAKTRERERQARERDY 1080
OY 1081 AEIODFHTFEGCDELMYGVSVSYEGSMALNARPOSPREGHMDALYAAYKPRNSKSP 1140
DB 1081 AEIODFHTFEGCDELMYGVSVSYEGSMALNARPOSPREGHMDALYAAYKPRNSKSP 1140
OY 1141 VDSNRSTPSNNDRIQRLROEFOAKODEYEDRRRTYSFEQOPWPNARPATOGGRISYVE 1200
DB 1141 VDSNRSTPSNNDRIQRLROEFOAKODEYEDRRRTYSFEQOPWPNARPATOGGRISYVE 1200
OY 1201 VQMKROQOEERESSQAQROYSSLPROSKNMASSVSDSWEDQNTSPGEGFOSAKENPRTS 1260

```

```

DB 1198 VQMKROQOEERESSQAQROYSSLPROSKNMASSVSDSWEDQNTSPGEGFOSAKENPRTS 1257
OY 1261 SYGSRNGYIGGFGFNARVYLETOELLROBORRQEOAKKOPSEGPSNYSYKRVODPS 1320
DB 1258 SYGSRNGYIGGFGFNARVYLETOELLROBORRQEOAKKOPSEGPSNYSYKRVODPS 1317
OY 1321 YAPPKGFRODVPSPSOVARLRLQTPPEKGRPFS 1356
DB 1318 YAPPKGFRODVPSPSOVARLRLQTPPEKGRPFS 1353

```

```

RESULT 2
O9B57 PRELIMINARY; PRT: 1031 AA.
AC O9B57;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
OS ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN SHORT VARIANT.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fang C., Xu Y.;
RT "Exon/Intron Structure and Splicing Variants of a Novel Human Polarity
RT Gene, hASIP.";
RL Submitted (CCIT-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: A0196186; AAK27892.1;
DR HSPG; Q12923; 3PDZ;
DR InterPro: IPR001478; PDZ;
DR Pfam: PF00595; PDZ; 3;
DR SMART: SM00228; PDZ; 3;
DR PROSITE: PS50106; PDZ; 3;
SQ SEQUENCE 1031 AA; 113418 MW; 40DD593636AE999 CRC64;

```

Query Match Best Local Similarity 57.6%; Score 781; DB 4; Length 1031; Matches 1021; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

```

OY 1 MKVYCFGRTRVYVPCGGHMKVSLIQAVTRYKRAIADPNWYLIQVHLEHGGDGIID 60
DB 1 MKVYCFGRTRVYVPCGGHMKVSLIQAVTRYKRAIADPNWYLIQVHLEHGGDGIID 60
OY 61 LDDILCDVADKDRLVAVFDEQDPHHGGDGTSASTGTOSPELFGSELGTNNVSAFOPYQ 120
DB 61 LDDILCDVADKDRLVAVFDEQDPHHGGDGTSASTGTOSPELFGSELGTNNVSAFOPYQ 120
OY 121 ATSELEVTPSVLRANMPLHVRSSDPALIGLSTVSQSNFSESPSKNPTRMSTTAGFL 180
DB 121 ATSELEVTPSVLRANMPLHVRSSDPALIGLSTVSQSNFSESPSKNPTRMSTTAGFL 180
OY 181 KONTAGSPPTCKDKKDEYRSLPRDTSNWSNOFORNARSLASASHPMVKMLEKOEODE 240
DB 181 KONTAGSPPTCKDKKDEYRSLPRDTSNWSNOFORNARSLASASHPMVKMLEKOEODE 240
OY 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGLGIHVVPFSAARGRTLG 300
DB 241 DGTEDNSRVEPVGHADTGLEHINPNSLDDMYKLYEVPNNGGLGIHVVPFSAARGRTLG 300
OY 301 LLYKLEKGGKAHEHENLFRENDCLVRINDGLNRRREFEQAHMFROAMRTPIIMFHVPA 360
DB 301 LLYKLEKGGKAHEHENLFRENDCLVRINDGLNRRREFEQAHMFROAMRTPIIMFHVPA 360
OY 361 ANKEQYEQLSQSEKNYSSRSPDSQYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHR 420
DB 361 ANKEQYEQLSQSEKNYSSRSPDSQYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHR 420
OY 421 LPHSAHPGKPPASAPASQNVSTVSSGYNTKKIGKRLNQLKKGTEGLGFSTSRDV 480

```

Db 421 LPHSAHPGKPPAPASAPQNVSTTVSSGYNTKTKIGKRLNIOCLKGTEGLGFSITSRDV 480
Qy 481 TIGGSAPLYVKNILPRGAALIDGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGTV 540
Db 481 TIGGSAPLYVKNILPRGAALIDGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGTV 540
Qy 541 SILVFRQEDAFHPRRLNAEPQOMQIPKTKADEDEDIVLPDGTREFLTFEVPPLNDGSGAG 600
Db 541 SILVFRQEDAFHPRRLNAEPQOMQIPKTKADEDEDIVLPDGTREFLTFEVPPLNDGSGAG 600
Qy 601 LCVSVKGNKSKNHADLGIFFVKSIIINGGAASKDGRLVNDQLLAVNGESLLGTNDADAE 660
Db 601 LCVSVKGNKSKNHADLGIFFVKSIIINGGAASKDGRLVNDQLLAVNGESLLGTNDADAE 660
Qy 661 TLRSKSTGKNGMQLIVARRISKCNELKSPGSPGPELPIETALDRERRISHSLS 720
Db 661 TLRSKSTGKNGMQLIVARRISKCNELKSPGSPGPELPIETALDRERRISHSLS 720
Qy 721 GIEGLDESPSRNAALSRINGESGKYQLSPTVNMPODDTVIIEDDRLLPVLPHLSGSSSS 780
Db 721 GIEGLDESPSRNAALSRIM---GKYQLSPTVNMPODDTVIIEDDRLLPVLPHLSGSSSS 777
Qy 781 SHDDVGFVYADAGTAAKAIISDASCSLSPDVDPVLAFOREGGRGSMSEKRTKOFSDAS 840
Db 778 SHDDVGFVYADAGTAAKAIISDASCSLSPDVDPVLAFOREGGRGSMSEKRTKOFSDAS 837
Qy 841 QLDFFVTKRSKMDLGIADETKLNVDQKAGSPSRDVGPSLGKSSLSLSTQTAFAV 900
Db 838 QLDFFVTKRSKMDLGIADETKLNVDQKAGSPSRDVGPSLGKSSLSLSTQTAFAV 897
Qy 901 TLNGDIPFHRPRRIIRGRGKNESEFRAAIDKSYDKPAVDDEGEMTLEEDTESSRSGR 960
Db 898 TLNGDIPFHRPRRIIRGRGKNESEFRAAIDKSYDKPAVDDEGEMTLEEDTESSRSGR 957
Qy 961 EESVSTASDQPSLSLEONMNGEKDXTDRKDKTGKERRKDRKDKKAKKMLKGL 1020
Db 958 EESVSTASDQPSLSLEONMNGEKDXTDRKDKTGKERRKDRKDKKAKKMLKGL 1017
Qy 1021 GDMF 1024
Db 1018 GDMF 1021

RESULT 3
Q96RM7 PRELIMINARY: PRT: 988 AA.
AC 096RM7:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN SHORT VARIANT
DE B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pang C., Xu Y.;
RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity
Gene, HASIP.";
RL EMBL: AF332593; AAK69192.1; -;
DR EMBL: AF332593; AAK69192.1; -;
SQ SEQUENCE 988 AA; 108545 MW; 89F2139B096F7E7F CMC64;

Query Match 41.0%; Score 556; DB 4; Length 988;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKVTGEGRTRVVPCGDMKVFSLIOQAVTRYRAIAKADPYWIOVHRLHGGDGLD 60
Db 1 MKVTGEGRTRVVPCGDMKVFSLIOQAVTRYRAIAKADPYWIOVHRLHGGDGLD 60

Qy 61 LDDILCDVADDKDRLVAVEDEDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120
Db 61 LDDILCDVADDKDRLVAVEDEDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120
Qy 121 ATSEIEVTSVLANPPLHARRSSDPALIGLSTSVSDSNFSSSEPSKAPTRKSTTAGFL 180
Db 121 ATSEIEVTSVLANPPLHARRSSDPALIGLSTSVSDSNFSSSEPSKAPTRKSTTAGFL 180
Qy 181 KONTAGSPKTCDDKKNENRSLPRDTSNMSNOQORANRSSLASHPMVKMLEKODE 240
Db 181 KONTAGSPKTCDDKKNENRSLPRDTSNMSNOQORANRSSLASHPMVKMLEKODE 240
Qy 241 DGTEDNSRVEPYGHADTGLEHINPFLSDMVAFLVEPNNGGGLGIHVVPFARGRTLG 300
Db 241 DGTEDNSRVEPYGHADTGLEHINPFLSDMVAFLVEPNNGGGLGIHVVPFARGRTLG 300
Qy 301 LLYKLEKGGKAEHMLFRENDICIVRINDGDLNRRFEQAQHFROAMRTPIIFHVPA 360
Db 301 LLYKLEKGGKAEHMLFRENDICIVRINDGDLNRRFEQAQHFROAMRTPIIFHVPA 360
Qy 361 ANKEQYEOLOSSEKNNYSSRSPDSQYIDNRSVNSAGLHTVQAPRLNHPPOIOSHR 420
Db 361 ANKEQYEOLOSSEKNNYSSRSPDSQYIDNRSVNSAGLHTVQAPRLNHPPOIOSHR 420
Qy 421 LPHSAHPGKPPAPASAPQNVSTTVSSGYNTKTKIGKRLNIOCLKGTEGLGFSITSRDV 480
Db 421 LPHSAHPGKPPAPASAPQNVSTTVSSGYNTKTKIGKRLNIOCLKGTEGLGFSITSRDV 480
Qy 481 TIGGSAPLYVKNILPRGAALIDGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGTV 540
Db 481 TIGGSAPLYVKNILPRGAALIDGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGTV 540
Qy 541 SILVFRQEDAFHPRREL 556
Db 541 SILVFRQEDAFHPRREL 556

RESULT 4
Q96RM6 PRELIMINARY: PRT: 1273 AA.
AC 096RM6:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN LONG VARIANT
DE B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pang C., Xu Y.;
RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity
Gene, HASIP.";
RL EMBL: AF332593; AAK69193.1; -;
DR EMBL: AF332593; AAK69193.1; -;
SQ SEQUENCE 1273 AA; 141730 MW; 53C1A94D8CB7341E CMC64;

Query Match 41.0%; Score 556; DB 4; Length 1273;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKVTGEGRTRVVPCGDMKVFSLIOQAVTRYRAIAKADPYWIOVHRLHGGDGLD 60
Db 1 MKVTGEGRTRVVPCGDMKVFSLIOQAVTRYRAIAKADPYWIOVHRLHGGDGLD 60
Qy 61 LDDILCDVADDKDRLVAVEDEDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120
Db 61 LDDILCDVADDKDRLVAVEDEDPHHGGDGTSSSTGTOSPEIFGSELGTNNVSAPQPY 120

```
QY 121 ANSEIVETSVLRANMPLHVRRSSDPALIGLSTVSQSNFSSSEPSRKNPTTRWSTTAGFL 180
DB 121 ANSEIVETSVLRANMPLHVRRSSDPALIGLSTVSQSNFSSSEPSRKNPTTRWSTTAGFL 180
QY 181 KONTASPTCKCKKCKKENTRSLPRTOSNWSNQFORNASSLSASHPMYGKMLEKODE 240
DB 181 KONTASPTCKCKKCKKENTRSLPRTOSNWSNQFORNASSLSASHPMYGKMLEKODE 240
QY 241 DGTEDNSVEPEVGHADTLEHLPNFSLDDMYLVEVPNDGPGLGTHVPEFSAARGRTLG 300
DB 241 DGTEDNSVEPEVGHADTLEHLPNFSLDDMYLVEVPNDGPGLGTHVPEFSAARGRTLG 300
QY 301 LUYKLEKCKGKAHEMLFPENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPA 360
DB 301 LUYKLEKCKGKAHEMLFPENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPA 360
QY 361 ANKEOYEOLOSSEKKNYSRFSPODYIDNRSVNSAGLHTVORARLHNPPEQIDSHSR 420
DB 361 ANKEOYEOLOSSEKKNYSRFSPODYIDNRSVNSAGLHTVORARLHNPPEQIDSHSR 420
QY 421 LPHSAHPSGKPPSAPASAPQNFSTTVSSGYNTKKIKGRLNIQKKTEGLGFSITSRDY 480
DB 421 LPHSAHPSGKPPSAPASAPQNFSTTVSSGYNTKKIKGRLNIQKKTEGLGFSITSRDY 480
QY 481 TTGGSPPIYVKNILPRGAIIDGRLKAGDRLIEVNGVDLVGKSQSEBVSLRSTMEGTIV 540
DB 481 TTGGSPPIYVKNILPRGAIIDGRLKAGDRLIEVNGVDLVGKSQSEBVSLRSTMEGTIV 540
QY 541 SILVFRQEDAFHPRRL 556
DB 541 SILVFRQEDAFHPRRL 556

RESULT 5
Q9NVE6 PRELIMINARY: PRT: 1266 AA.
ID AC Q9NVE6:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PAR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=20394296; PubMed=10934474;
RA Joberly G., Petersen C., Gao L., Macara I.G.;
RT "The cell-polarity protein Par6 links Par3 and atypical protein kinase
RT C to Cdc42."
RL Nat. Cell Biol. 2:531-539(2000).
DR EMBL: AF252293; AAF1530.1; -.
DR HSSP: Q12923; 3PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PS50106; PDZ; 3.
SQ SEQUENCE 1266 AA; 141071 MW; 92DF51B68081AA42 CRC64;
```

Query Match 36.9%; Score 500; DB 4; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 857 IADFKLTWVDQKAGSRVGSGLGKSSLSLQTAVAEVLNGDIPFHRPRRII 916
DB 767 IADFKLTWVDQKAGSRVGSGLGKSSLSLQTAVAEVLNGDIPFHRPRRII 826
QY 917 RGRGNESFRAIDKSYKPAVDDDEGMETLEEDTSSSGRESVSTAADOSSHSLR 976
```

```
DB 827 RGRGNESFRAIDKSYDPAVDDDEGMETLEEDTSSSGRESVSTAADOSSHSLR 886
QY 977 QMNGOKEKDKTDKKDKTGKCKKDRDKEKDKAKKGLGLDMFRFGKRRDKIE 1036
DB 887 QMNGOKEKDKTDKKDKTGKCKKDRDKEKDKAKKGLGLDMFRFGKRRDKIE 946
QY 1037 KKGKIKIOESFSEBERIKMOEERIOAKTREFREROAREVDVAEIDPFHTFGCDEL 1096
DB 947 KKGKIKIOESFSEBERIKMOEERIOAKTREFREROAREVDVAEIDPFHTFGCDEL 1006
QY 1097 MTGVSVSEGSMAIARQSPREGHMDALYAQVKPRNKRSPVDSNRSTPSNDRTOR 1156
DB 1007 MTGVSVSEGSMAIARQSPREGHMDALYAQVKPRNKRSPVDSNRSTPSNDRTOR 1066
QY 1157 LKQEOQAKODEVDDBRRRTYSFEQPMNARATQSGHSVVEVQOMORQOEFESSQ 1216
DB 1067 LKQEOQAKODEVDDBRRRTYSFEQPMNARATQSGHSVVEVQOMORQOEFESSQ 1126
QY 1217 AQOYRSLPQSRKNASSVSDSWQNTSPGEGFQSAKENPRYSQSGRNGYLGGHFN 1276
DB 1127 AQOYRSLPQSRKNASSVSDSWQNTSPGEGFQSAKENPRYSQSGRNGYLGGHFN 1186
QY 1277 ARVMELELLRQEOQRKEQKKKOPSEGPSNDYKRVQDPSYAPKGFRODVPSP 1336
DB 1187 ARVMELELLRQEOQRKEQKKKOPSEGPSNDYKRVQDPSYAPKGFRODVPSP 1246
QY 1337 SOVARLNRLQTPKGRPPYS 1356
DB 1247 SOVARLNRLQTPKGRPPYS 1266
```

```
RESULT 6
Q9HC48 PRELIMINARY: PRT: 667 AA.
ID AC Q9HC48:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CTCL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=2143360; PubMed=11149944;
RA Eichmuller D., Usener D., Dummer R., Stein A., Thiel D.,
RA Schandendorf S.;
RT "Serological detection of cutaneous T-cell lymphoma-associated
RT antigens."
RL Proc. Natl. Acad. Sci. U S A. 98:629-634(2001).
DR EMBL: AF177228; AAC33676.1; -.
DR HSSP: Q12923; 3PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 2.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 667 AA; 73499 MW; C653BC16802B4B02 CRC64;
```

Query Match 30.5%; Score 413; DB 4; Length 667;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 513; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 313 EHENLFRENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPAANKEOYEOLOS 372
DB 1 EHENLFRENDCIYRINDGLRNRRFEQAQHMFRQAMPTPIIWEHVPAANKEOYEOLOS 60
QY 373 EKNNYSSRFSPODYIDNRSVNSAGLHTVORARLHNPPEQIDSHSRLLPHSAHPSGPP 432
```

```

Db 61 EKNNTSSRSPDSQYIDNRSVNSAGLHTVORAPRLNHPEQIDSHSLRPHSAHPSGKPP 120
OY 433 SAPASAPQNVFTTSSVSGYNTKIGKRLNLTQKKTGEGLFSTTSRDVITGSGAPITYKN 492
Db 121 SAPASAPQNVFTTSSVSGYNTKIGKRLNLTQKKTGEGLFSTTSRDVITGSGAPITYKN 180
OY 493 ILPRGAIIDGRLKAGDRLIEVNGVDLVGKSQEEVVSLLRSTMEGTVSLVFRQEDAFH 552
Db 181 ILPRGAIIDGRLKAGDRLIEVNGVDLVGKSQEEVVSLLRSTMEGTVSLVFRQEDAFH 240
OY 553 PRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPLNDGSGAGVSKGNSRKE 612
Db 241 PRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPLNDGSGAGVSKGNSRKE 300
OY 613 NHADLGIPEKSIINGGAASKDGRRLRVNDOLIAVNGESLLGKTQNMAMETLLRSMTEGKN 672
Db 301 NHADLGIPEKSIINGGAASKDGRRLRVNDOLIAVNGESLLGKTQNMAMETLLRSMTEGKN 360
OY 673 RGMIOQLIVARRISKCNELKSPGSPGPPELPIETALDDRERRISHSLYSGIEGLDESPSRN 732
Db 361 RGMIOQLIVARRISKCNELKSPGSPGPPELPIETALDDRERRISHSLYSGIEGLDESPSRN 420
OY 733 AALSRIMGSSGKYQLSPTVNMPODDVITIEDRLPVLPHLSPOSSSSSHDDVGFVTADA 792
Db 421 AALSRIMGSSGKYQLSPTVNMPODDVITIEDRLPVLPHLSPOSSSSSHDDVGFVTADA 480
OY 793 GTWAKAALSDSADCSLSPDVDPVLAFORSGFGHQ 826
Db 481 GTWAKAALSDSADCSLSPDVDPVLAFORSGFGHQ 514

```

```

RESULT 7
ID 096K28 PRELIMINARY; PRT; 865 AA.
AC 096K28:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE CDNA FLJ14829 FIS, CLONE OVARC1000945, MODERATELY SIMILAR TO
DE RATUUS NORREGICUS ATYPICAL PKC SPECIFIC BINDING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OVARIAN CARCINOMA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H.,
RA Wagahtsuma M., Hosofiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murekami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027735; BAB55330.1; -.
SQ SEQUENCE 865 AA; 95161 MW; 5378BBCD406D0835 CRC64;

```

```

Query Match 29.9%; Score 406; DB 4; Length 865;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 234 EKQEDDEDETEEDNSRVEPRGHADTGLEHLPNFSLDQVLYVEVPDNGGLGIHVVPESA 233
Db 99 EKQEDDEDETEEDNSRVEPRGHADTGLEHLPNFSLDQVLYVEVPDNGGLGIHVVPESA 158
OY 294 RGGRTLGLLVKRLKKGAKAHEMLFRENDCIYRINDGDLNRRPEQAQAHMFQAMRTPII 353
Db 159 RGGRTLGLLVKRLKKGAKAHEMLFRENDCIYRINDGDLNRRPEQAQAHMFQAMRTPII 218
OY 354 WFIHVPAANKQEYEQLSQSEKNNYSSRSPDSQYIDNRSVNSAGLHTVORAPRLNHPE 413

```

```

Db 219 WFIHVPAANKQEYEQLSQSEKNNYSSRSPDSQYIDNRSVNSAGLHTVORAPRLNHPE 278
OY 414 QIDSHSRRLPSHAPSGKPPASAPAPQNVFTTSSVSGYNTKIGKRLNLTQKKTGEGLF 473
Db 279 QIDSHSRRLPSHAPSGKPPASAPAPQNVFTTSSVSGYNTKIGKRLNLTQKKTGEGLF 338
OY 474 SITSRDVITGSGAPITYKNLIPRGAAIODGRKLAGDRLIEVNGVDLVGKSQEEVVSLLRS 533
Db 339 SITSRDVITGSGAPITYKNLIPRGAAIODGRKLAGDRLIEVNGVDLVGKSQEEVVSLLRS 398
OY 534 TKMEGTVSLVFRQEDAFHPRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPL 593
Db 399 TKMEGTVSLVFRQEDAFHPRELNAEPSOMQIPRETKAEDEDIVLPDGTREFLFEVPL 458
OY 594 NDSGAGLVSVGNRKENHADIGIVKSTINGGAASKDGRRLRVNDOLIAVNGESLLGK 653
Db 459 NDSGAGLVSVGNRKENHADIGIVKSTINGGAASKDGRRLRVNDOLIAVNGESLLGK 518
OY 654 TNODAMETLLRSMTEGKNRGMIOQLIVARRISKCNELKSPGSPGPPELPIETALDDRERR 713
Db 519 TNODAMETLLRSMTEGKNRGMIOQLIVARRISKCNELKSPGSPGPPELPIETALDDRERR 578
OY 714 ISHSLYSGIEGLDESPSRNALSRIMG 740
Db 579 ISHSLYSGIEGLDESPSRNALSRIMG 605

```

```

RESULT 8
ID 09NML4 PRELIMINARY; PRT; 347 AA.
AC 09NML4:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ20754 FIS, CLONE HEP02246 (UNKNOWN) (PROTEIN FOR
DE MGC:19518).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000761; BAA91366.1; -.
DR EMBL; BC011711; AAH11711.1; -.
SQ SEQUENCE 347 AA; 40538 MW; BE2B3557996EC91E CRC64;

```

```

Query Match 25.6%; Score 347; DB 4; Length 347;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1010 MKKAKGMLKGLGMFRFGKRRKDDIKETGKIRKIOESFTSEERIRMKQOERIQAOTRE 1069
Db 1 MKKAKGMLKGLGMFRFGKRRKDDIKETGKIRKIOESFTSEERIRMKQOERIQAOTRE 60
OY 1070 FREQOARERDYAEIODEFHRFFGCDDDELMTYGVSSEGSMLNARPOSREGHMMDALYAO 1129
Db 61 FREQOARERDYAEIODEFHRFFGCDDDELMTYGVSSEGSMLNARPOSREGHMMDALYAO 120
OY 1130 VKKPRNSKPSPVDSNRSTPSNHDRIQRLROFOAKODEVEDERRRITYSEQPPNAPRA 1189
Db 121 VKKPRNSKPSPVDSNRSTPSNHDRIQRLROFOAKODEVEDERRRITYSEQPPNAPRA 180

```

QY 1190 TOSGRHSYVEVOMQROREBRESSOAOROYSLPROSRRKNASSVSODSWEONYSPGEG 1249
 DB 181 TOSGRHSYVEVOMQROREBRESSOAOROYSLPROSRRKNASSVSODSWEONYSPGEG 240
 QY 1250 FOSAKENPRYSYOGSNGTIGSGHNARVMELETOELLROBORKEOAKKOPPSGSPN 1309
 DB 241 FOSAKENPRYSYOGSNGTIGSGHNARVMELETOELLROBORKEOAKKOPPSGSPN 300
 QY 1310 YDSYKRVODPSYAPKGPPODVPSPSOVARLNRLQTPKGRPFYS 1356
 DB 301 YDSYKRVODPSYAPKGPPODVPSPSOVARLNRLQTPKGRPFYS 347

RESULT 9
 ID Q92340 PRELIMINARY; PRT; 1337 AA.
 AC Q92340;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DE ATYPICAL PKC SPECIFIC BINDING PROTEIN.
 GN ASP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98437350; PubMed=9763423;
 RA Izumi Y., Hirose T., Tamai Y., Hirai S., Nagashima Y., Fujimoto T.,
 RA Tabuse Y., Kempfues K.J., Ohno S.;
 RT "An atypical PKC directly associates and colocalizes at the epithelial
 RT tight junction with ASP, a mammalian homologue of caenorhabditis
 RL U. Cell Biol. 143:95-106(1998).
 DR EMBL; AB005549; BAA34216.1;
 DR HSBP; Q12923; 3PDZ.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ.
 DR SMART: SM00228; PDZ.
 DR PROSITE: PS50106; PDZ.
 SQ SEQUENCE 1337 AA; 149448 MW; EC980C5106B52F9C CRC64;

Query Match
 Best Local Similarity 6.5%; Score 88; DB 11; Length 1337;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 566 PRETAEDEDIVLPDGTREFLFEVPLNDGSGAGVSVKGNRKENHADLGIFFVKSII 625
 DB 566 PRETAEDEDIVLPDGTREFLFEVPLNDGSGAGVSVKGNRKENHADLGIFFVKSII 625
 QY 626 NGGAASKDGRLRVNDOLIAVNGESLIGK 653
 DB 626 NGGAASKDGRLRVNDOLIAVNGESLIGK 653
 RESULT 10
 ID Q99NH2 PRELIMINARY; PRT; 1333 AA.
 AC Q99NH2;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DE PAR-3 180 KDA ISOFORM.
 GN PAR3 OR PAR3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-NIH SWISS;
 RX MEDLINE=99121117; PubMed=9920925;
 RA Lin D., Gish G.D., Songyang Z., Pawson T.;
 RT "The carboxyl terminus of B class ephrins constitutes a PDZ domain
 RT binding motif".
 RL J. Biol. Chem. 274:3726-3733(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NIH SWISS;
 RX MEDLINE=20394297; PubMed=10934475;
 RA Lin D., Edwards A.S., Fawcett J.P., Mbamalu G., Scott J.D., Pawson T.;
 RT "A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC
 RL signalling and cell polarity".
 RL Nat. Cell Biol. 2:540-547(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NIH SWISS;
 RA Lin D.C.J.; Pawson T.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY026057; AAK07669.1;
 DR HSBP; Q12923; 3PDZ.
 DR MGI; 2135608; Pard3.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ.
 DR SMART: SM00228; PDZ.
 DR PROSITE: PS50106; PDZ.
 SQ SEQUENCE 1333 AA; 149060 MW; AF67825C66DCFE86 CRC64;

Query Match
 Best Local Similarity 5.7%; Score 77; DB 11; Length 1333;
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 577 VLPDGTREFLFEVPLNDGSGAGVSVKGNRKENHADLGIFFVKSIIINGGAASKDGR 636
 DB 577 VLPDGTREFLFEVPLNDGSGAGVSVKGNRKENHADLGIFFVKSIIINGGAASKDGR 636
 QY 637 RVNDOLIAVNGESLIGK 653
 DB 637 RVNDOLIAVNGESLIGK 653

RESULT 11
 ID Q96DK9 PRELIMINARY; PRT; 624 AA.
 AC Q96DK9;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DE CDNA FL25236 F15, CLONE STM02096 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=GASTRIC MUCOSA;
 RA Nimmiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 RA Nakagawa K., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,
 RA Irie K., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
 RA Nagai K., Isogai T., Sugano S.;
 RT NEDO human cDNA sequencing project.
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK057965; BAB71623.1;
 DR NON_TER 624
 FT 624
 SQ SEQUENCE 624 AA; 68002 MW; 749EC5B79F79F096 CRC64;

Query Match
 Best Local Similarity 1.4%; Score 19; DB 4; Length 624;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 882 LGKKSSLESLOTAAVEV 900
 DB 462 LGKKSSLESLOTAAVEV 480

RESULT 12

OY 096NX6 PRELIMINARY; PRT; 1143 AA.
 AC 096NX6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICING VARIANT B.
 GN PARL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gao L., Macara I.G., Joberty G.;
 RT "Multiple splice variants of Par3 and of a novel related gene, Par3L,
 RT produce functionally different proteins."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF428251; AAL3065.1; -
 SQ SEQUENCE 1143 AA; 126102 MW; 368B69CF81D45E7E CRC64;

Query Match 1.4%; Score 19; DB 4; Length 1143;
 Best Local Similarity 100.0%; Pred. No. 1.3e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 882 LGKKSSLESLOTAAVEV 900
 DB 677 LGKKSSLESLOTAAVEV 695

RESULT 13

OY 096NX7 PRELIMINARY; PRT; 1205 AA.
 AC 096NX7;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICING VARIANT A.
 GN PARL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gao L., Macara I.G., Joberty G.;
 RT "Multiple splice variants of Par3 and of a novel related gene, Par3L,
 RT produce functionally different proteins."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF428250; AAL3066.1; -
 SQ SEQUENCE 1205 AA; 132494 MW; 26E6704CCDCE8C8 CRC64;

Query Match 1.4%; Score 19; DB 4; Length 1205;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 882 LGKKSSLESLOTAAVEV 900
 DB 739 LGKKSSLESLOTAAVEV 757

RESULT 14

OY 096N09 PRELIMINARY; PRT; 545 AA.

AC 096N09;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CDNA FLJ131595 FIS. CLONE NT2R12002517, WEAKLY SIMILAR TO HOMO SAPIENS
 DE PAR3 mRNA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nidomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuo Y., Nagai K., Isogai T.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK056157; BAB71106.1; -
 SQ SEQUENCE 545 AA; 58691 MW; 03013B66B510FF0 CRC64;

Query Match 1.0%; Score 13; DB 4; Length 545;
 Best Local Similarity 100.0%; Pred. No. 0.00019;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 592 PLNDGSGAGIGVS 604
 DB 500 PLNDGSGAGIGVS 512

RESULT 15

OY 095LT2 PRELIMINARY; PRT; 128 AA.
 AC 095LT2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOPHYSAL 14.4 KDA PROTEIN.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopitheidae; Macaca.
 OX NCBI_Taxid=9341;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
 RA Teruo K., Sugano S.;
 RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
 RT libraries."
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB071110; BAB64504.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 128 AA; 14436 MW; AF92C43CC505F93F CRC64;

Query Match 0.9%; Score 12; DB 6; Length 128;
 Best Local Similarity 100.0%; Pred. No. 0.00056;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1325 KGPRQDVPPSP 1336
 DB 75 KGPRQDVPPSP 86

Search completed: July 24, 2002, 14:37:02
 Job time: 208 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:32:04 ; Search time 18.11 Seconds
(without alignments)
1828.885 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356
Sequence: 1 MKVTVCGRKRRVVPCGDGH.....SQVARLNRLQTPKGRPEYS 1356Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	0.6	31	3	US-09-045-632-70
2	8	0.6	31	3	US-09-045-632-73
3	8	0.6	79	3	US-09-100-804-26
4	8	0.6	297	2	US-09-151-611-3
5	8	0.6	297	4	US-09-370-102-3
6	8	0.6	348	3	US-08-415-655-5
7	8	0.6	348	3	US-08-415-655-13
8	8	0.6	348	3	US-08-415-655-15
9	8	0.6	388	1	US-08-445-640-6
10	8	0.6	388	3	US-08-170-558-6
11	8	0.6	388	3	US-08-447-314-6
12	8	0.6	388	3	US-08-445-461-6
13	8	0.6	450	2	US-08-665-037-2
14	8	0.6	450	2	US-08-666-067-2
15	8	0.6	450	2	US-08-732-870-2
16	8	0.6	890	1	US-08-445-640-2
17	8	0.6	890	3	US-08-170-558-2
18	8	0.6	890	3	US-08-447-314-2
19	8	0.6	890	3	US-08-445-461-2
20	8	0.6	911	1	US-08-286-305A-1
21	8	0.6	911	2	US-08-441-104A-1
22	8	0.6	911	2	US-08-440-816A-1
23	8	0.6	911	4	US-09-417-381A-1
24	8	0.5	61	4	US-09-314-268-145
25	8	0.5	100	3	US-08-851-843A-10
26	7	0.5	100	4	US-08-974-549A-192
27	7	0.5	100	4	US-08-854-050-10

28	7	0.5	100	4	US-09-430-323-10	Sequence 10, Appl
29	7	0.5	128	4	US-09-199-637A-79	Sequence 79, Appl
30	7	0.5	182	1	US-08-345-756-2	Sequence 2, Appl1
31	7	0.5	182	1	US-08-625-198-2	Sequence 2, Appl1
32	7	0.5	305	3	US-08-965-600-1	Sequence 1, Appl1
33	7	0.5	320	2	US-08-800-264A-6	Sequence 6, Appl1
34	7	0.5	320	2	US-09-018-628-6	Sequence 6, Appl1
35	7	0.5	320	3	US-09-273-378-6	Sequence 6, Appl1
36	7	0.5	320	3	US-09-209-605-6	Sequence 6, Appl1
37	7	0.5	333	3	US-08-966-118-3	Sequence 3, Appl1
38	7	0.5	353	4	US-09-216-619-3	Sequence 3, Appl1
39	7	0.5	377	3	US-08-888-077A-27	Sequence 27, Appl
40	7	0.5	384	3	US-08-946-026-27	Sequence 27, Appl
41	7	0.5	437	2	US-08-800-264A-7	Sequence 7, Appl1
42	7	0.5	437	2	US-09-018-628-7	Sequence 7, Appl1
43	7	0.5	437	3	US-09-273-378-7	Sequence 7, Appl1
44	7	0.5	437	3	US-09-209-605-7	Sequence 7, Appl1
45	7	0.5	439	1	US-08-519-103-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-045-632-70
; Sequence 70, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, Richard L.
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corliss, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-045-632-70

Query Match 0.6%; Score 8; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 DGRLRVND 640
Db 15 DGRLRVND 22

RESULT 2

US-09-045-632-73
; Sequence 73, Application US/09045632
; Patent No. 6001575
; GENERAL INFORMATION:
; APPLICANT: HUGANIR, Richard L.
; APPLICANT: Dong, Hualing
; TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
; TITLE OF INVENTION: GRIP-RELATED MOLECULES
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/045,632
; FILING DATE: 19-MAR-1998
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,016
; FILING DATE: 19-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48147/1699-CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-045-632-73

Query Match 0.6%; Score 8; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 DGRLRVND 640
Db 15 DGRLRVND 22

RESULT 3

US-09-100-804-26
; Sequence 26, Application US/09100804
; Patent No. 6066472
; GENERAL INFORMATION:
; APPLICANT: GONEZ, LEONEL JORGE
; APPLICANT: SARAS, JAN
; APPLICANT: CLAESSEN-WELSH, LENA
; APPLICANT: HELDIN, CARL-HENRIK
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,804
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/596,291
; FILING DATE: 09-AUG-1996
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09943
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: 10461/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYDROTHERMAL: NO
; ANTI-SENSE: NO
; US-09-100-804-26

Query Match 0.6%; Score 8; DB 3; Length 79;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 633 DGRLRVND 640
Db 39 DGRLRVND 46

RESULT 4

US-09-151-611-3
; Sequence 3, Application US/09151611
; Patent No. 5958731
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Au-Young, Janice
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
; FILE REFERENCE: PR-0599 US
; CURRENT APPLICATION NUMBER: US/09/151,611
; CURRENT FILING DATE: 1998-09-11
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE: -
; OTHER INFORMATION: 91685067

US-09-151-611-3

Query Match 0.6%; Score 8; DB 2; Length 297;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 640 DOLIAVNG 647
|||||
DB 245 DOLIAVNG 252

RESULT 5

US-09-370-102-3
; Sequence 3, Application US/09370102
; Patent No. 6265547
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Au-Young, Janice
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
; FILE REFERENCE: PF-0599 US
; CURRENT APPLICATION NUMBER: US/09/370,102
; EARLIER APPLICATION NUMBER: 09/151,611
; EARLIER FILING DATE: 1998-09-11
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE: -
; OTHER INFORMATION: g1685067
US-09-370-102-3

Query Match 0.6%; Score 8; DB 4; Length 297;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 640 DOLIAVNG 647
|||||
DB 245 DOLIAVNG 252

RESULT 6

US-08-415-655-5
; Sequence 5, Application US/08415655
; Patent No. 6025480
; GENERAL INFORMATION:
; APPLICANT: Massague, Joan
; APPLICANT: Lee, Mong-hong
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: P75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF
; TITLE OF INVENTION: SAME
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,655
; FILING DATE:
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1747/47418
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-415-655-5

Query Match 0.6%; Score 8; DB 3; Length 348;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438
|||||
DB 139 PPSAPASA 146

RESULT 7

US-08-415-655-13
; Sequence 13, Application US/08415655
; Patent No. 6025480
; GENERAL INFORMATION:
; APPLICANT: Massague, Joan
; APPLICANT: Lee, Mong-hong
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: P75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF
; TITLE OF INVENTION: SAME
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,655
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1747/47418
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-415-655-13

Thu Jul 25 08:38:17 2002

us-09-757-781-2.olg.ra1

Query Match 0.6%; Score 8; DB 3; Length 348;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438
DB 139 PPSAPASA 146

RESULT 8
US-08-415-655-15
Sequence 15, Application US/08415655
Patent No. 6025480
GENERAL INFORMATION:
APPLICANT: Massague, Joan
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: P75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF
TITLE OF INVENTION: SAME
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,655
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1747/47418
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ. ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-415-655-15

Query Match 0.6%; Score 8; DB 3; Length 348;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438
DB 139 PPSAPASA 146

RESULT 9
US-08-445-640-6
Sequence 6, Application US/08445640
Patent No. 5709858
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases

NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,640
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ. ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-640-6

Query Match 0.6%; Score 8; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 434 APASAPON 441
DB 281 APASAPON 288

RESULT 10
US-08-170-558-6
Sequence 6, Application US/08170558
Patent No. 6001621
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESS: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/170,558
FILING DATE: 20-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-170-558-6

Query Match
Best Local Similarity 100.0%; Score 8; DB 3; Length 388;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 APASAPON 441
DB 281 APASAPON 288

RESULT 11
US-08-447-314-6
Sequence 6, Application US/08447314
Patent No. 6087144
GENERAL INFORMATION:
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,314
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-447-314-6

Query Match
Best Local Similarity 100.0%; Score 8; DB 3; Length 388;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 APASAPON 441
DB 281 APASAPON 288

RESULT 12
US-08-445-461-6
Sequence 6, Application US/08445461
Patent No. 6096527
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/445,461
FILING DATE: 22-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-445-461-6

Query Match
Best Local Similarity 100.0%; Score 8; DB 3; Length 388;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 APASAPON 441
DB 281 APASAPON 288

RESULT 13

US-08-665-037-2
Sequence 2, Application US/08665037
Patent No. 5895813
GENERAL INFORMATION:
APPLICANT: Seedorf, Klaus
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
OF TRKA-1 RELATED
TITLE OF INVENTION: DISORDERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,037
FILING DATE: June 13, 1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/005,167
FILING DATE: October 13, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 220/156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-665-037-2

Query Match

Best Local Similarity 0.6%; Score 8; DB 2; Length 450;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 DRLIEVNG 516
|||||||

DB 193 DRLIEVNG 200

RESULT 14

US-08-666-067-2
Sequence 2, Application US/08666067
Patent No. 5922842
GENERAL INFORMATION:
APPLICANT: Seedorf, Klaus
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
OF TRKA-1 RELATED
TITLE OF INVENTION: DISORDERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street

STREET: Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB

COMPUTER: IBM Compatible

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/666,067

FILING DATE: June 13, 1996

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/005,421

FILING DATE: October 13, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 220/157

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 450 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-666-067-2

Query Match

Best Local Similarity 0.6%; Score 8; DB 2; Length 450;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 DRLIEVNG 516
|||||||

DB 193 DRLIEVNG 200

RESULT 15

US-08-732-870-2
Sequence 2, Application US/08732870
Patent No. 5945523
GENERAL INFORMATION:
APPLICANT: Seedorf, Iulgard
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
OF TRKA-1 RELATED
TITLE OF INVENTION: DISORDERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/732,870
FILING DATE: October 15, 1996
CLASSIFICATION: 536

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/666,037
;; FILING DATE: June 13, 1996
;; APPLICATION NUMBER: 08/666,067
;; FILING DATE: June 13, 1996
;; APPLICATION NUMBER: 60/005,167
;; FILING DATE: October 13, 1995
;; APPLICATION NUMBER: 60/005,421
;; FILING DATE: October 13, 1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warburg, Richard J.
;; REGISTRATION NUMBER: 32,327
;; REFERENCE/DOCKET NUMBER: 222/247
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 955-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ. ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 450 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-732-870-2

Query Match 0.6%; Score 8; DB 2; Length 450;
Best local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 509 DRLIEVNG 516
|||
Db 193 DRLIEVNG 200

Search completed: July 24, 2002, 14:35:23
Job time: 199 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:34:09 ; Search time 16.09 Seconds

(without alignments)
3263.127 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356
Sequence: 1 MKYTVCFGRTRVVVPCGDH.....SQVARNLRLOTPEKRPFTS 1356

Scoring table: OLIGO
Gapop 60.0 , Capext 60.0

Searched: 105224 seqs, 38719550 residues

Word size: 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	0.7	1693	Y163_SYNY3	05563 synechocyst
2	8	0.6	200	P6AD_BACFI	P94300 bacillus fi
3	8	0.6	283	P635_METUA	Q58052 methanococ
4	8	0.6	348	CDNC_MOUSE	P49919 mus musculu
5	8	0.6	475	Z131_HUMAN	P52739 homo sapien
6	8	0.6	488	VAB1_HORVU	Q40078 hordeum vul
7	8	0.6	592	YK06_CAEEL	P34294 caenorhabdi
8	8	0.6	724	DLG4_MOUSE	Q62108 mus musculu
9	8	0.6	724	DLG4_RAT	P31016 rattus norv
10	8	0.6	737	YNC2_CAEEL	P34535 caenorhabdi
11	8	0.6	767	DLG4_HUMAN	P78352 homo sapien
12	8	0.6	852	DLG2_RAT	Q63622 rattus norv
13	8	0.6	870	DLG2_HUMAN	Q15700 homo sapien
14	8	0.6	890	TYO3_HUMAN	Q06418 homo sapien
15	8	0.6	904	DLG1_HUMAN	Q12959 homo sapien
16	8	0.6	911	DLG1_RAT	Q62996 rattus norv
17	8	0.6	1081	SY1_TERTH	P36422 tetrahymena
18	8	0.6	1490	CRK7_HUMAN	Q9nyv4 homo sapien
19	8	0.6	4036	RRPL_DUGBV	Q66431 dupre virus
20	7	0.5	79	Y476_ARCFU	Q29774 archaeoglob
21	7	0.5	107	KVID_HUMAN	P01596 homo sapien
22	7	0.5	116	Y789_METUA	Q58199 methanococ
23	7	0.5	142	YOHN_BACSU	P54512 bacillus su
24	7	0.5	150	VE6_HPV42	P27229 human papil
25	7	0.5	177	RL10_MYCLE	Q9cbk7 mycobacteri
26	7	0.5	178	RL10_MYCTU	P96940 mycobacteri
27	7	0.5	180	YRBI_HAEIN	P45314 haemophilus
28	7	0.5	181	OM20_NEUCR	P35848 neuropsora
29	7	0.5	246	YIAT_ECO57	P58242 escherichia
30	7	0.5	246	YIAT_ECOLI	P37681 escherichia
31	7	0.5	248	TPM3_CHICK	P19353 gallus gall
32	7	0.5	258	YNM8_YEAST	P53862 saccharomyc
33	7	0.5	261	MOTB_BACSU	P28612 bacillus su

34	7	0.5	270	1	VA32_VARV	P33849 variola vir
35	7	0.5	276	1	ATP6_SYNY3	P27178 synechocyst
36	7	0.5	287	1	T215_ARATH	O23403 arabidopsis
37	7	0.5	297	1	YNFL_ECOLI	P75559 escherichia
38	7	0.5	300	1	VA32_VACCV	P21055 vaccinia vl
39	7	0.5	312	1	BLAB_STRCI	P33652 streptomyc
40	7	0.5	342	1	EP1X_CLOPE	P18020 clostridium
41	7	0.5	354	1	LPXD_CHLMU	Q9pkf1 chlamydia m
42	7	0.5	354	1	RECA_SYNY3	P74737 synechocyst
43	7	0.5	371	1	OP25_HAEIN	P46027 haemophilus
44	7	0.5	376	1	PSD4_MOUSE	Q35226 mus musculu
45	7	0.5	377	1	NIV1_ANASP	Q44290 anabaena sp

ALIGNMENTS

```
RESULT 1
Y163_SYNY3 STANDARD; PRT; 1693 AA.
AC 05563;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Hypothetical WD-repeat protein SL0163.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Keneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 644 to 924 of the genome."
RL DNA Res. 2:153-166(1995).
CC 1- SIMILARITY: CONTAINS 16 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: D63999; BAA10064.1; -
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 16.
CC PRINTS: PR00320; GPROTEINRPT.
CC SMART: SM00320; WD40; 16.
CC PROSITE: PS00678; WD_REPEATS_1; 8.
CC PROSITE: PS50082; WD_REPEATS_2; 15.
CC PROSITE: PS50294; WD_REPEATS_REGION; 1.
CC Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1008 1042
FT REPEAT 1053 1083
FT REPEAT 1094 1124
FT REPEAT 1135 1165
FT REPEAT 1176 1206
FT REPEAT 1217 1247
FT REPEAT 1258 1288
FT REPEAT 1299 1329
FT REPEAT 1340 1370
FT REPEAT 1381 1411
FT REPEAT 1422 1452
FT REPEAT 1463 1493
FT REPEAT 1504 1534
FT REPEAT 1545 1575
FT REPEAT 1586 1616
FT REPEAT 1627 1657
```

SO SEQUENCE 1693 AA; 189935 MW; 0977A827A0251CFP CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 1; Length 1693;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 381 RFPSPSOYI 389
1431 RFPSPSOYI 1439

RESULT 2

PAAD_BACFI
ID PAAD_BACFI STANDARD; PRT; 200 AA.
AC P94300;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
OS Probable aromatic acid decarboxylase (EC 4.1.1.-).
OC Bacillus firmus.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1399;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OF4;
RA Ito M., Kruwisch T.A.;
RT Cloning and sequence of gerc locus from alkaliphilic Bacillus
RT firmus OF4.
RL Submitted (JUN-1996) to the EMBL/Genbank/DDBJ databases.
CC -1 SIMILARITY: BELONGS TO THE POLYPEPTYL P-HYDROXYBENZOATE /
CC PHENYLACETIC ACID DECARBOXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U61168; AAB41845.1; -
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
DR Hypothetical protein; Iyase; Decarboxylase.
SQ SEQUENCE 200 AA; 22233 MW; 0C212E8AD11AC13D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 200;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1283 TOELLROE 1290
26 TOELLROE 33

RESULT 3

ID Y635_METHA STANDARD; PRT; 283 AA.
AC Q58052;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DT Hypothetical protein M0655.
GN M0655.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RA MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fumman J.L., Nguyen D.,
RA Cotton M.D., Roberts K.M., Hurs M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U67511; AAB98633.1; -
DR TIGR; M0635;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 283 AA; 32502 MW; 1CE52FC8457C5E20 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 283;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 999 KKKDRDKE 1006
245 KKKDRDKE 252

RESULT 4

ID CDNC_MOUSE STANDARD; PRT; 348 AA.
AC P49919;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Cyclin-dependent kinase inhibitor 1C (Cyclin-dependent kinase
DE inhibitor p57) (P57KIP2).
GN CDKN1C OR KIP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=95247027; PubMed=7729683;
RA Lee M.-H., Reynolds J.L., Massague J.;
RT "Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique
RT domain structure and tissue distribution."
RL Genes Dev. 9:639-649(1995).
RN [2]
RP SEQUENCE OF 14-348 FROM N.A.
RX MEDLINE=95247028; PubMed=7729684;
RA Matsuno S., Edwards M.C., Bai C., Parker S., Zhang P., Baldini A.,
RA Harper J.W., Elledge S.J.;
RT "p57KIP2, a structurally distinct member of the p21CIP1 Cdk inhibitor
RT family, is a candidate tumor suppressor gene."
RL Genes Dev. 9:650-662(1995).
CC -1- FUNCTION: POTENT TIGHT-BINDING INHIBITOR OF SEVERAL G1 CYCLIN/CDK
CC COMPLEXES (CYCLIN E-CDK2, CYCLIN D2-CDK4, AND CYCLIN A-CDK2) AND,
CC TO LESSER EXTENT, OF THE MITOTIC CYCLIN B-CDK2. NEGATIVE REGULATOR
CC OF CELL PROLIFERATION. MAY PLAY A ROLE IN MAINTENANCE OF THE
CC NONPROLIFERATIVE STATE THROUGHOUT LIFE.
CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- TISSUE SPECIFICITY: IT IS EXPRESSED IN THE HEART, BRAIN, LUNG,
CC SKELETAL MUSCLE, KIDNEY, PANCREAS AND TESTIS. HIGH LEVELS ARE SEEN
CC IN THE PLACENTA WHILE LOW LEVELS ARE SEEN IN THE LIVER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U20553; AAC52186.1; -
CC EMBL: U22399; AAA85096.1; -
CC HSSP: P46527; IJ5U.
CC MGD: MGI:104564; Cdkn1c.
CC InterPro: IPR003175; CDI.
CC Pfam: PF02234; CDI.1.
CC Cell cycle; Alternative splicing.
CC FT DOMAIN 108 189 PRO-RICH.
CC FT DOMAIN 178 284 GLU/ASP-RICH.
CC FT DOMAIN 309 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT VARSPLIC 1 13 MISSING (IN ISOFORM KIP2B/P57B).
CC FT CONFLICT 150 151 DA -> EP (IN REF. 2).
CC SQ SEQUENCE 348 AA; 37331 MW; 108A8538D77016D9 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 431 PPSAPASA 438
Db 139 PPSAPASA 146

DR MIM: 604073; -
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; zf-C2H2; 5.
DR SMART: SM00355; ZNF_C2H2; 5.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 5.
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW Nuclear protein; Repeat.
FT NON_TER 1 1
FT DOMAIN 113 295 ZINC FINGERS.
FT ZN_FING 113 135 C2H2-TYPE.
FT ZN_FING 140 163 C2H2-TYPE.
FT ZN_FING 180 202 C2H2-TYPE.
FT ZN_FING 244 266 C2H2-TYPE.
FT ZN_FING 272 295 C2H2-TYPE.
SQ SEQUENCE 475 AA; 54087 MW; 3388E158CC567FF CRC64;

Query Match 0.6%; Score 8; DB 1; Length 475;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1208 QERRESSQ 1215
Db 426 QERRESSQ 433

RESULT 5
Z131_HUMAN STANDARD; PRT; 475 AA.
AC P52739;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 131 (Fragment).
GN ZNF131.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Insulinoma;
RX MEDLINE=96044430; PubMed=7557990;
RA Tommerup N., Vissing H.;
RT "Isolation and fine mapping of 16 novel human zinc finger-encoding
RT cDNAs identify putative candidate genes for developmental and
RT malignant disorders.";
RT Genomics 27:259-264(1995).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U09410; AAC50251.1; -
CC HSSP: P07248; ZADR.

RESULT 6
VAB1_HORVU STANDARD; PRT; 488 AA.
ID VAB1_HORVU
AC Q40078;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vacuolar ATP synthase subunit B isoform 1 (EC 3.6.3.14) (V-ATPase B
DE subunit 1) (Vacuolar proton pump B subunit 1).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Root;
RX MEDLINE=94159793; PubMed=8115549;
RA Berkelman T., Houtchens K.A., Dupont F.M.;
RT "Two cDNA clones encoding isoforms of the B subunit of the vacuolar
RT ATPase from barley roots.";
RT Plant Physiol. 104:287-288(1994).
CC -1- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L11862; AAA81330.1; -
CC InterPro: IPR004100; ATP-synt_ab_N.
CC InterPro: IPR000793; ATPase_AB_C.
CC InterPro: IPR000194; ATPase_alpha_beta.
CC Pfam: PF00006; ATP-synt_ab; 1.

DR Pfam: PF00306; ATP-synt_ab_C: 1.
 DR Pfam: PF02874; ATP-synt_ab_N: 1.
 DR PROSITE: PS00152; ATPASE_ALPHA_BETA: 1.
 KW Hydrolyase; ATP synthetase; Hydrogen ion transport; Multigene family.
 SQ SOURCE 488 AA; 54026 MW; 0571B898CECC1070 CRC64;

Query Match 0.6%; Score 8; DB 1; Length 488;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 302 LVKRLKRG 309
 |||||
 DB 186 LVKRLKRG 193

RESULT 7
 YK06_CAEEL STANDARD; PRT; 592 AA.
 AC P34294;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 66.9 kDa protein C05B5.6 in chromosome III.
 GN C05B5.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Mortimore B.J.;
 RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME, TO C.ELEGANS ZK1290.9.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z32679; CA83594.1; -
 DR PIR: S43570; S43570.
 DR WormPep: C05B5.6; CE00052.
 DR InterPro: IPR002900; D0F38.
 DR Pfam: PF01827; D0F38; 1.
 KW Hypothetical protein.
 FT DOMAIN 9 15 POLY-SER.
 FT DOMAIN 448 455 POLY-LEU.
 FT DOMAIN 584 587 POLY-SER.
 SQ SEQUENCE 592 AA; 66879 MW; A0BCA915649710BD CRC64;

Query Match 0.6%; Score 8; DB 1; Length 592;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 774 SDQSSSS 781
 |||||
 DB 6 SDQSSSS 13

RESULT 8
 DLG4_MOUSE STANDARD; PRT; 724 AA.
 ID DLG4_MOUSE
 AC Q62108;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)

DE (Synapse-associated protein 90) (Discs, large homolog 4).
 GN DLG4 OR DLGH4 OR PSD95.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DRA/2; TISSUE-Brain;
 RA Kohmura N., Yagi T.;
 RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOSOLIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
 CC ON THE PRESYNAPTIC SIDE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D50621; BAA09297.1; -
 DR HSSP: P31016; IBE9.
 DR MGD: MGI:1277959; DLgh4.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 KW SH3 domain; Repeat.
 FT DOMAIN 65 151 PDZ 1.
 FT DOMAIN 160 246 PDZ 2.
 FT DOMAIN 313 393 PDZ 3.
 FT DOMAIN 428 498 SH3.
 FT DOMAIN 534 724 GUANYLATE_KINASE.
 SQ SEQUENCE 724 AA; 80472 MW; 7EFC99E1FF90BA CRC64;

Query Match 0.6%; Score 8; DB 1; Length 724;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 633 DGRLRVND 640
 |||||
 DB 108 DGRLRVND 115

RESULT 9
 DLG4_RAT STANDARD; PRT; 724 AA.
 ID DLG4_RAT
 AC P31016; P97631;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
 DE (Synapse-associated protein 90) (Discs, large homolog 4).
 GN DLG4 OR PSD95.
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;
 RX MEDLINE=93040233; PubMed=1419001;
 RA Cho K.-O., Hunt C.A., Kennedy M.B.;
 RT "The rat brain postsynaptic density fraction contains a homolog of
 RT the Drosophila discs-large tumor suppressor protein.";
 RL Neuron 9:929-942(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;
 RX MEDLINE=93186749; PubMed=7680343;
 RA Kistner U., Wenzel B.M., Voh R.W., Cases-Ianghoff C., Garner A.M.,
 RA Appeltner U., Voss B., Gundelfinger E.D., Garner C.;
 RT "SNAP90, a rat presynaptic protein related to the product of the
 RT Drosophila tumor suppressor gene *dlg-A*.";
 RL J. Biol. Chem. 268:4580-4583(1993).
 RN [3]
 RN SEQUENCE OF 566-625 FROM N.A.
 RC STRAIN-WISTAR KYOTO; TISSUE=Vascular smooth muscle;
 RA Adams L.D., Werny I., Schwartz S.M.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS) OF 302-402.
 RX MEDLINE=96270509; PubMed=8674113;
 RA Doyle D.A., Lee A., Lewis J., Kim E., Sheng M., Mackinnon R.;
 RT "Crystal structures of a complexed and peptide-free membrane protein-
 RT binding domain: molecular basis of peptide recognition by PDZ.";
 RL Cell 85:1067-1076(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
 CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
 CC POSTSYNAPTIC).
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M96853; AAA41971.1; -
 DR EMBL: X66474; CAA47103.1; -
 DR EMBL: U77090; AAB38270.1; -
 DR PIR: S26407; S26407.
 DR PIR: JH0800; JH0800.
 DR PDB: 1BE9; 21-OCY-98.
 DR PDB: 1BE9; 21-OCY-98.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00625; Guanylate_kin. 1.
 DR Pfam: PF00595; PDZ. 3.
 DR Pfam: PF00018; SH3. 1.
 DR SMART: SM00072; GUKC. 1.
 DR SMART: SM00228; PDZ. 3.
 DR SMART: SM00228; PDZ. 3.
 DR SMART: SM00326; SH3. 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS0052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS0106; PDZ. 3.
 DR PROSITE: PS0002; SH3. 1.
 DR SH3 domain; Repeat; 3D-structure.
 FT DOMAIN 65 151 PDZ 1.

FT DOMAIN 160 246 PDZ 2.
 FT DOMAIN 313 393 PDZ 3.
 FT DOMAIN 428 498 SH3.
 FT DOMAIN 534 724 GUANYLATE KINASE.
 FT CONFLICT 61 61 M -> L (IN REF. 2).
 FT CONFLICT 78 78 S -> T (IN REF. 2).
 FT CONFLICT 177 182 GGNQH -> ALGTS (IN REF. 2).
 FT CONFLICT 200 200 A -> G (IN REF. 2).
 FT CONFLICT 254 254 S -> T (IN REF. 2).
 FT CONFLICT 539 555 ILGPRKDRANDDLSEF -> ISLPKRTVPTMISPPSS
 (IN REF. 2).
 FT CONFLICT 623 625 (GRH -> RQD (IN REF. 3).
 SQ SEQUENCE 724 AA; 80465 MW; 7922D4BE0F9AD85 CRC64;
 Query Match 0.6%; Score 8; DB 1; Length 724;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 633 DGRLRVND 640
 Db 108 DGRLRVND 115
 RESULT 10
 YNC2_CAEEL
 ID YNC2_CAEEL STANDARD; PRT; 737 AA.
 AC P34535;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 83.6 kDa protein R05D3.2 in chromosome III.
 GN R05D3.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Cronin M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Garner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohlschlag P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: I07144; AK21441.1; -
 DR PIR: S44862; S44862.
 DR WormPep: R05D3.2; CE00281.
 KW Hypothetical protein.
 SQ SEQUENCE 737 AA; 83555 MW; 3397543C5C6EC9B4 CRC64;
 Query Match 0.6%; Score 8; DB 1; Length 737;
 Best Local Similarity 100.0%; Pred. No. 17;

Thu Jul 25 08:38:25 2002

```

Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

OY      376  NYSRRFS 383
        |||||||
Db       453  NYSRRFS 460

RESULT  11
DIG4_HUMAN STANDARD; PRT; 767 AA.
ID      DIG4_HUMAN
AC      P78352: 092941: Q9UKK8;
DT      01-NOV-1997 (Rel. 35, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Presynaptic density protein 95 (PSD-95) (Discs, large homolog 4).
OS      Homo sapiens (human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_Taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Mammary gland;
RA      MEDLINE=97432823; PubMed=9286702;
RT      Stathakis D.G., Hoover K.B., You Z., Bryant P.J.;
RT      "Human postsynaptic density-95 (PSD95): location of the gene (DLG4)
RT      and possible function in nonneural as well as in neural tissues.";
RT      genomics 44:71-82(1997).
RN      [2]
RP      REVISIONS.
RC      TISSUE=Mammary gland;
RA      Stathakis D.G., Hoover K.H., You Z., Bryant P.J.;
RL      Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=20047407; PubMed=10582582;
RA      Stathakis D.G., Udar N., Sandgren O., Andreasson S., Bryant P.J.;
RA      Small K., Forstman-Semk K.;
RT      "Genomic organization of human DLG4, the gene encoding postsynaptic
RT      density 95.";
RT      J. Neurochem. 73:2250-2265(1999).
RN      [4]
RP      SEQUENCE OF 81-401 FROM N.A.
RC      TISSUE=Brain;
RA      Bremner J.E., Parkinson J.F., Manzano W.P., McClary J.A.;
RL      Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: INTERACTS WITH THE CYTOSOLIC TAIL OF NMDA RECEPTOR
CC      SUBUNIT. MAY BE INVOLVED IN SYNAPTIC JUNCTIONS PRIMARILY
CC      -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS TO BE
CC      ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
CC      POSTSYNAPTIC).
CC      -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
CC      -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC      -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC      -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC      -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
CC      -----
CC      EMBL: 083192; AAC52113.1;
CC      EMBL: AF156495; AAD56173.1;
CC      EMBL: U68138; AAB07736.1;
CC      HSSP: P31016; IBE9.
CC      MIM: 602887;
CC      InterPro: IPR000619; Guanylate_kin.
CC      InterPro: IPR001478; PDZ.
CC      InterPro: IPR001452; SH3.

```

```

DR      Pfam; PF00625; Guanylate_kin; 1.
DR      Pfam; PF00595; PDZ; 3.
DR      Pfam; PF00018; SH3; 1.
DR      SMART; SM0072; GUKC; 1.
DR      SMART; SM00228; PDZ; 3.
DR      SMART; SM00326; SH3; 1.
DR      PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR      PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR      PROSITE; PS50106; PDZ; 3.
DR      PROSITE; PS50002; SH3; 1.
DR      SH3 domain; Repeat.
KW      SH3 domain.
FT      DOMAIN 108 194 PDZ 1.
FT      DOMAIN 203 289 PDZ 2.
FT      DOMAIN 356 436 PDZ 3.
FT      DOMAIN 471 561 SH3.
FT      DOMAIN 577 767 GUANYLATE_KINASE.
FT      DOMAIN 46 46 E->V (IN REF. 3).
FT      CONFLICT 81 83 VIV->V (IN REF. 4).
FT      CONFLICT 399 401 GDO->AGI (IN REF. 4).
FT      CONFLICT 399 401 GDO->AGI (IN REF. 4).
FT      SEQUENCE 767 AA; 85429 MW; BE1019159E5B2D8 CRC64;

Query Match      0.6%; Score 8; DB 1; Length 767;
Best Local Similarity 100.0%; Pred. No. 18;
Matches      8;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

OY      633  DGLRYND 640
        |||||||
Db       151  DGLRYND 158

RESULT  12
DIG2_RAT STANDARD; PRT; 852 AA.
ID      DIG2_RAT
AC      Q63622; Q62939; P70548;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Channel associated protein of synapse-110 (Synaptic
DE      density protein PSD-93) (Discs, large homolog 2).
GN      Dlg2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_Taxid=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=96310881; PubMed=8755482;
RA      Kim E., Cho K.O., Rothschild A., Sheng M.;
RT      "Heteromultimerization and NMDA receptor-clustering activity of
RT      Chapsyn-110, a member of the PSD-95 family of proteins.";
RT      Neuron 17:103-113(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=96193770; PubMed=8625413;
RA      Bremner J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,
RA      Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,
RA      Froehner S.C., Bredt D.S.;
RT      "Interaction of nitric oxide synthase with the postsynaptic density
RT      protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
RT      Cell 84:757-767(1996).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RA      Irie M., Hata Y., Takai Y.;
RL      FUNCTION: INTERACTS WITH THE CYTOSOLIC TAIL OF NMDA RECEPTOR
CC      -1- FUNCTION: AS WELL AS POTASSIUM CHANNELS.
CC      SUBUNIT: AS WELL AS POTASSIUM CHANNELS.
CC      -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC      -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC      -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC      -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U49049; AAB53243.1; -
 DR EMBL: U50717; AAC52643.1; -
 DR EMBL: U53368; AAB48562.1; -
 DR HSSP: O12959; 1PDR.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 KM SH3 domain: Repeat.
 FT DOMAIN 98 184 PDZ 1.
 FT 193 279 PDZ 2.
 FT DOMAIN 421 501 PDZ 3.
 FT 536 606 SH3.
 FT DOMAIN 662 852 GUANYLATE_KINASE.
 FT 181 182 VR -> IL (IN REF. 2).
 FT 228 228 I -> M (IN REF. 2).
 FT 228 228 R -> K (IN REF. 2).
 FT 339 339 D -> E (IN REF. 3).
 FT 450 454 MISSING (IN REF. 2).
 FT 464 465 GD -> RK (IN REF. 2).
 FT 474 474 D -> H (IN REF. 2).
 FT 476 476 R -> P (IN REF. 2).
 FT 478 478 A -> D (IN REF. 2).
 FT 486 486 AAA -> LP (IN REF. 2).
 FT 506 506 A -> S (IN REF. 2).
 FT 569 569 H -> N (IN REF. 2).
 FT 586 586 L -> Q (IN REF. 2).
 FT 641 641 MISSING (IN REF. 2).
 FT 639 639 K -> A (IN REF. 3).
 FT 726 726 F -> L (IN REF. 1).
 FT 733 733 N -> Y (IN REF. 2).
 FT 749 749 E -> V (IN REF. 1).
 FT 756 756 L -> H (IN REF. 2).
 FT 791 792 KR -> NG (IN REF. 2).
 FT 794 794 T -> M (IN REF. 2).
 SQ SEQUENCE 852 AA; 94934 MW; F8D41AAB9C5B09 CRC64;
 Query Match 0.6%; Score 8; DB 1; Length 852;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 633 DGRRLRVND 640
 DB 141 DGRRLRVND 148
 RESULT 13
 ID DLG2_HUMAN STANDARD; PRT; 870 AA.
 AC 015700;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Channel associated protein of synapse-110 (Chapsyn-110) (Discs, large
 DE homolog 2).
 GN DLG2.

OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96310881; PubMed=8755482;
 RA Kim E., Cho K.-O., Rothschild A., Sheng M.;
 RT "Heteromultimerization and NMDA receptor-clustering activity of
 RT Chapsyn-110, a member of the PSD-95 family of proteins.";
 RL Neuron 17:103-113(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNIT AS WELL AS POTASSIUM CHANNELS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32376; AAB04949.1; -
 DR HSSP: O12959; 1PDR.
 DR MIM: 603583; -
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 KM SH3 domain: Repeat.
 FT DOMAIN 98 184 PDZ 1.
 FT 193 279 PDZ 2.
 FT DOMAIN 421 501 PDZ 3.
 FT 536 606 SH3.
 FT DOMAIN 680 870 GUANYLATE_KINASE.
 SQ SEQUENCE 870 AA; 97500 MW; 89C83BA0619F6F59 CRC64;
 Query Match 0.6%; Score 8; DB 1; Length 870;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 633 DGRRLRVND 640
 DB 141 DGRRLRVND 148
 RESULT 14
 ID TYO3_HUMAN STANDARD; PRT; 890 AA.
 AC 006418; 014953;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Tyrosine-protein kinase receptor TYRO3 precursor (EC 2.7.1.112)
 DE (Tyrosine-protein kinase RSE) (Tyrosine-protein kinase SKY) (Tyrosine-
 DE protein kinase DTK) (Protein-tyrosine kinase BYK).
 GN TYRO3 OR RSE OR SKY OR DTK OR BYK.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Carnivora; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94193774; PubMed=7511603;
 RT Mark M.R., Scadden D.T., Wang Z., Gu Q., Goddard A., Godowski P.J.;
 RT "RSE, a novel receptor-type tyrosine kinase with homology to Axl/Ufo,
 RT is expressed at high levels in the brain.";
 RT J. Biol. Chem. 269:10720-10728(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94150991; PubMed=8108112;
 RT Chashi K., Mizuno K., Kuma K., Miyata T., Nakamura T.;
 RT "Cloning of the cDNA for a novel receptor tyrosine kinase, Sky,
 RT predominantly expressed in brain.";
 RT Oncogene 9:699-705(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95161080; PubMed=7857658;
 RT Crosier K.E., Hall L.R., Lewis P.M., Morris C.M., Wood C.R.,
 RT "Isolation and characterization of the human DTK receptor tyrosine
 RT kinase.";
 RT Growth Factors 11:137-144(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97318332; PubMed=9175267;
 RT Kaji Y., Nishimura Y., Kato M., Mizuguchi M., Saji M., Katsunuma T.,
 RT Ohno K., Takashima S., Onodera K.;
 RT "A tyrosine kinase-like molecule is localized in the nuclear membrane
 RT of neurons: hippocampal behavior under stress.";
 RT Biol. Cell 88:45-54(1996).
 RN [5]
 RP SEQUENCE OF 519-720 FROM N.A.
 RA MEDLINE=94085793; PubMed=8262388;
 RT Polvi A., Armstrong E., Lal C., Lemke G., Huebner K., Spritz R.A.,
 RT Gluda L.C., Nicholls R.D., Allitalo R.;
 RT "The human TYRO3 gene and pseudogene are located in chromosome
 RT 1q24-q25.";
 RN [6]
 RP GENE 134:289-293(1993).
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL ADHESION PROCESSES, PARTICULARLY
 CC IN THE CENTRAL NERVOUS SYSTEM.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ABUNDANT IN THE BRAIN AND LOWER LEVELS IN
 CC OTHER TISSUES.
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 005682; AAA19236.1;
 DR EMBL: D17517; BAA04467.1;
 DR EMBL: 018934; AAC50070.1;
 DR EMBL: D50479; BAA21781.1;
 DR EMBL: X72886; CAA51396.1;
 DR HSSP: P11362; IRGK.
 DR MIM: 600341;
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.

DR InterPro: IPR003600; Ig_Like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00041; fn3_2.
 DR Pfam: PF00047; Ig_2.
 DR Pfam: PF00069; pkinase_1.
 DR SMART: SM00060; FN3_2.
 DR SMART: SM00410; Ig_Like_1.
 DR SMART: SM00408; Ig2_1.
 DR SMART: SM00219; Tyrc_1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM_1.
 DR KMW: KMW001; Glycoprotein: Tyrosine-protein kinase: ATP-binding;
 DR KMW: KMW001; Phosphorylation: Transmembrane; Signal; Repeat;
 DR KMW: KMW001; Immunoglobulin domain.
 DR FT SIGNAL 1 40
 DR FT CHAIN 1 40
 DR FT DOMAIN 41 890
 DR FT TRANSMEM 41 429
 DR FT DOMAIN 430 450
 DR FT DOMAIN 451 890
 DR FT DOMAIN 57 124
 DR FT DOMAIN 153 210
 DR FT DOMAIN 224 311
 DR FT DOMAIN 315 411
 DR FT DOMAIN 518 790
 DR FT NP_BIND 524 532
 DR FT BINDING 550 550
 DR FT ACT_SITE 550 550
 DR FT BINDING 655 655
 DR FT DISULFID 64 117
 DR FT DISULFID 160 203
 DR FT CARBOHYD 63 63
 DR FT CARBOHYD 191 191
 DR FT CARBOHYD 230 230
 DR FT CARBOHYD 240 240
 DR FT CARBOHYD 293 293
 DR FT CARBOHYD 366 366
 DR FT CARBOHYD 380 380
 DR FT MOD_RES 686 686
 DR FT CONFLICT 29 36
 DR FT CONFLICT 285 285
 DR FT CONFLICT 293 293
 DR FT CONFLICT 341 341
 DR FT CONFLICT 812 812
 DR FT CONFLICT 815 815
 DR FT CONFLICT 890 AA; 96505 MM; F9EC675077C4B8F1 CRC64;
 DR SEQUENCE

Query Match
 Best Local Similarity 0.6%; Score 8; DB 1; Length 890;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 434 APASAPON 441
 DB 321 APASAPON 328

RESULT 15
 D1G1_HUMAN
 ID D1G1_HUMAN STANDARD; PRT; 904 AA.
 AC 012959; 012958;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, last annotation update)
 DE Presynaptic protein SAP97 (Synapse-associated protein 97) (Discs,
 DE large homolog 1).
 GN D1G1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Carnivora; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95024052; PubMed=7937897;

```

RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
RT "Cloning and characterization of hdlg; the human homologue of the
RT Drosophila discs large tumor suppressor binds to protein 4.1.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
RA MEDLINE:96338231; PUBMed:8757139;
RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
RA Marfatia S.M., Chishti A.H., Liddington R.C.;
RT "Crystal structure of a PDZ domain.";
RL Nature 382:649-652(1996).
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NR2A RECEPTOR
CC SUBUNITS. ASSOCIATES WITH PROTEIN 4.1.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U13897; AAA50599.1; -.
DR EMBL: U13896; AAA50598.1; -.
DR PDB: 1PDR; 23-JUL-97.
DR MIM: 601014; -.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00018; SH3; 1.
DR SMART: SM00072; GUKC; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain; Repeat; Alternative splicing; 3D-structure.
FT DOMAIN 224 310 PDZ 1.
FT DOMAIN 319 405 PDZ 2.
FT DOMAIN 466 546 PDZ 3.
FT DOMAIN 581 651 SH3.
FT DOMAIN 714 904 GUANYLATE KINASE.
FT VARSPPLIC 669 680 EIPDDMGSKGLK -> OSFNDKRRKRLFSRKFFPYKKNKDOS
FT EOETSDAQ (IN ISOFORM 2).
SQ SEQUENCE 904 AA; 100354 MW; B78798DBB0920D4 CRC64;

```

```

Query Match 0.6%; Score 8; DB 1; Length 904;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 633 DGRLRVND 640
DB 267 DGRLRVND 274

```

```

Search completed: July 24, 2002, 14:37:35
Job time: 206 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:30:49 ; Search time 30.55 Seconds

(without alignments)
4930.150 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 1356
Sequence: 1 MKYVCFGRIRVVVPCGDG.....SQVARNLRQLQPEKGRPFYS 1356

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq.032802:*

1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	406	29.9	865	22	AA895798 Human protein sequ
2	406	29.9	865	22	AA867610 Amino acid sequenc
3	195	14.4	272	22	AA674163 Human colon cancer
4	195	14.4	273	22	AAU17403 Novel signal trans
5	118	8.7	419	22	AB604716 Novel human diagno
6	110	8.1	744	21	AA771410 PDZ domain-compris
7	80	5.9	193	22	AB604714 Novel human diagno
8	60	4.4	60	22	AB643115 Peptide #10621 enc
9	60	4.4	60	22	AA836946 Peptide #10983 enc
10	60	4.4	90	21	AA771424 PDZ domain-compris
11	47	3.5	169	22	AB604715 Novel human diagno

12	38	2.8	51	21	AA802888 Human secreted pro
13	31	2.3	86	21	AA771423 PDZ domain-compris
14	11	0.8	187	22	AB604627 Novel human diagno
15	11	0.8	315	22	AB811149 Human PAR3 homolog
16	11	0.8	1464	22	AB611199 Drosophila melanog
17	8	0.6	41	21	AA853218 Protein containing
18	8	0.6	81	22	AA853384 Propionibacterium
19	8	0.6	106	22	AA853844 Propionibacterium
20	8	0.6	106	22	AA865767 Propionibacterium
21	8	0.6	128	21	AA809344 Arabidopsis thalia
22	8	0.6	135	21	AA838205 Human secreted pro
23	8	0.6	136	21	AA809343 Arabidopsis thalia
24	8	0.6	207	22	AA857625 Human homolog of D
25	8	0.6	207	22	AA858034 PDZ encoded domain
26	8	0.6	208	22	AA855833 PDZ encoded domain
27	8	0.6	211	22	AB610975 Wheat geminivirus
28	8	0.6	287	20	AA888295 Human prostate tum
29	8	0.6	303	20	AA774141 Human olfactory re
30	8	0.6	307	22	AA871576 Human OR-like poly
31	8	0.6	307	22	AA872477 Human post-synapti
32	8	0.6	344	22	AA855834 Human post-synapti
33	8	0.6	344	22	AA857623 Human cyclin-depe
34	8	0.6	344	22	AA858035 Murine cyclin-depe
35	8	0.6	348	17	AA801437 Human colon cancer
36	8	0.6	364	22	AA875048 Human developmenta
37	8	0.6	366	15	AA860550 Partial PSD-95 pro
38	8	0.6	396	18	AA834666 Mammalian two-hybr
39	8	0.6	414	22	AA836687 Tyrosine kinase as
40	8	0.6	450	19	AA846606 Human tyrosine kin
41	8	0.6	450	20	AA829635 Human tyrosine kin
42	8	0.6	450	20	AA822130 Zinc finger protei
43	8	0.6	510	22	AA844890 Drosophila melano
44	8	0.6	583	22	AB63033 Drosophila G-prote
45	8	0.6	583	22	AA838948

ALIGNMENTS

RESULT	ID	AA895798 standard; Protein: 865 AA.	ALIGNMENTS
XX	AA895798;		
XX	26-JUN-2001 (first entry)		
DE	Human protein sequence SEQ ID NO:18775.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy.		
OS	Homo sapiens.		
XX	EP1074617-A2.		
PN	07-FEB-2001.		
XX	28-JUL-2000; 2000EP-0116126.		
PF	29-JUL-1999; 99JP-0248036.		
XX	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX	(HELI-) HELIX RES INST.		
PA	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
PI	WPI; 2001-318749/34.		
DR	Primer sets for synthesizing polynucleotides, particularly the 5602		
XX			
PT			

full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 18775; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH01166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 865 AA:

Query Match 29.9%; Score 406; DB 22; Length 865;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

234 EKQDDEGDEEDNSRVEPGHADTGLHNPNSLDDMKLVLPNDGGLGIHVVPFSA 293
|||||
99 ekqgedegdeednsrvepghadtglhnpnslddmvklvepndggpiglihvvpfsa 158

294 RGGRTGLLVKRLKKGKAEHENLFRNDCTIVRINDGDLRNRREFQAOHMFROAMRPTII 353
|||||
159 rggrtglglvkrlekgykaehenlfrndctivrindgdlnrrefeqaohmfrqamrtptii 218

354 WFHVVPAAKQEQYBOLQSSEKNNTYSSRFPSDSQYIDNRSVNSAGLHTVORAPRLNHPPE 413
|||||
219 wfhvvpaaankeqeqylsqseknnyssrfpsdsqyidnrsvnsagltvorpaprlnhppe 278

414 QIDSHSRLLPHSAHPGSKPPSPAPASAPQNVSTVSSGYNTKTKGRRLNQLKGTBGLGF 473
|||||
279 qidshsrllphsaahpqsppspapasaqnvstsvssgynstkikgrlnqlkgtbglgf 338

474 SITSDVITIGSAPRIYVNNILPRGAATODGRUKAGDRLEIENGVDLVKSGSEEVSLIRS 533
|||||
339 sitsdvritigsapriyvnnilprgaatodgrukagdrleienvgdlvkgseeevslirs 398

534 TKMBEGVSLVLRQDAFHPRLEINAEPSQMOIPEKTKAEDBDIVLTPDGTREPTLFEVPL 593
|||||
399 tkmbegvslvlvrqdafhprelnaepsqmoipektkaedbdivltpdgtreptlfevpl 458

594 NDGSGNGLGVSVKGRNSKENHADLGIIFYKSTINGGAASKDGLRVRNDOLIVNGSSLIGK 653
|||||
459 ndgsgnlgvsvkgrnskenhadlgiifykslinggaaskdglrvrndolivngssligk 518

654 TNDQAEETLRSGMSTEGNKGMIQILIVARRISKCNELKSPGSPPELPIETALDDRERR 713
|||||
519 tndqaeetlrsgmstegnkgmiqilivarrikskcnelspgsppelpietalddrerr 578

714 ISHSLVSGIEGLDESPSRNAALSRLMG 740
|||||
579 ishslvsgiegldepsrnaalstrimg 605

RESULT 2

AA67610
ID AAG67610 standard; Protein; 865 AA.

AA67610;

26-NOV-2001 (first entry)

Amino acid sequence of a human protein kinase/protein phosphatase.

Human; protein kinase; protein phosphatase; signal transduction.

Homo sapiens.

NO200109316-A1.

08-FEB-2001.

28-JUL-2000; 2000MO-JP05061.

29-JUL-1999; 99JP-0248036.

18-OCT-1999; 99US-0159590.

11-JAN-2000; 2000JP-0118776.

17-FEB-2000; 2000US-0183322.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isegaki T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
Senoo C, Nezu J;

WPI; 2001-570286/64.

N-PSDB; AAH78369.

New genes encoding proteins with protein kinase/protein phosphatase activity, useful in the diagnosis and treatment of diseases -

Claim 2; Page 76-82; 233pp; Japanese.

The present sequence represents human protein kinase/protein phosphatase. It is expected that the protein kinase/protein phosphatase gene participates in signal transduction in cells. The protein kinase/protein phosphatase polypeptides and polynucleotides are useful for developing

phosphatase and treatment agents for human and animal diseases. The diagnosis and treatment agents for human and animal diseases. The

phosphatase polynucleotides are useful as target molecules in designing novel drugs. The protein kinase/protein

phosphatase polynucleotides are useful as a source of probes and primers, which may be used to isolate homologous sequences.

Sequence 865 AA:

Query Match 29.9%; Score 406; DB 22; Length 865;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 506; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

234 EKQDDEGDEEDNSRVEPGHADTGLHNPNSLDDMKLVLPNDGGLGIHVVPFSA 293
|||||
99 ekqgedegdeednsrvepghadtglhnpnslddmvklvepndggpiglihvvpfsa 158

294 RGGRTGLLVKRLKKGKAEHENLFRNDCTIVRINDGDLRNRREFQAOHMFROAMRPTII 353
|||||
159 rggrtglglvkrlekgykaehenlfrndctivrindgdlnrrefeqaohmfrqamrtptii 218

354 WFHVVPAAKQEQYBOLQSSEKNNTYSSRFPSDSQYIDNRSVNSAGLHTVORAPRLNHPPE 413
|||||
219 wfhvvpaaankeqeqylsqseknnyssrfpsdsqyidnrsvnsagltvorpaprlnhppe 278

414 QIDSHSRLLPHSAHPGSKPPSPAPASAPQNVSTVSSGYNTKTKGRRLNQLKGTBGLGF 473
|||||
279 qidshsrllphsaahpqsppspapasaqnvstsvssgynstkikgrlnqlkgtbglgf 338

QY 474 STSRDVTYIGGSAPIYVKNLPRGAALQDGRLAGDRLEIVNGVDLVGKSQEEVVSILRS 533
 |||||||
 Db 339 slsrdrvltlgssaspiyvknlprgaalqdgrrlkagdrllievngvdlvyksqeevvsllrs 398
 QY 534 TKMEGVSVLLVFPEQDAFHPRELNAEPSSQWQIKETKRAEDEDIVLPDGRREFLTFEVPVL 593
 |||||||
 Db 399 lkmegivsvllvtrqedaflhprelnaepsqmqipketaededlvpdgrrefllfevpl 458
 QY 594 NDGSGAGLVSVKGNRSKKNHADLGFVKSIINGAASKDRRLRVNDQLAVNGESILGK 653
 |||||||
 Db 459 sdgsagqlgysvkvgnrskenhadlglfvksllnggaaskqgrrlvndqlaavngesllgk 518
 QY 654 TNDAMETLRSMSTEGNKRGMQLIVARRISKCNELKSPGSPPELPETALDDRRER 713
 |||||||
 Db 519 lngdametlrrsmstegnkrgmqlivarriskcnelkspgsppepelpetaldrrer 578
 QY 714 ISHSLYSGIEGLDESPSRNALSRMG 740
 |||||||
 Db 579 lshslysglegldepsrnaalsrlmg 605

RESULT 3

ID AAG74163 standard; Protein; 272 AA.

AC AAG74163;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:4927.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 9905-0157137.

PR 03-NOV-1999; 9905-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI; 2001-235357/24.

DR N-PSDB; AAH33594.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

XX useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 6691-6693; 9803PP; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the

CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

SO Sequence 272 AA;

Query Match 14.4%; Score 195; DB 22; Length 272;
 Best Local Similarity 100.0%; Pred. No. 6.8e-194;
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 CIVRINDGLRNRRFEQAQHFQAMRPTIWFHVPAANEQYEQLSQSEKNYSSRF 382
 |||||||
 Db 65 clvrindgdlrnrrfegqahmfirgamtptliwfhvpaanekeqeqlsqseknysrfl 124
 QY 383 SPDSQYIDNRSVNSAGLHTVQRAERLNRHPPEQIDSHSLRPSAHPSCGRPSAPAPQNV 442
 |||||||
 Db 125 spdsqyidnrsvnsaglhtvqraprlnhppeqidsrhlpsahpsgkxpsapapnrv 184
 QY 443 FSTVSSGYNPKKIGKRLNIQLKKGTBGLGFSISRDVTIGGSAPIYVKNLPRGAALQD 502
 |||||||
 Db 185 fstvssgyntkkgikrlnqlkkgtbglgfsisrdvtlgssaspiyvknlprgaalqd 244
 QY 503 GRLKAGDRLEIVNGV 517
 |||||||
 Db 245 grlkagdrllievngv 259

RESULT 4

ID AAU17403 standard; Protein; 273 AA.

AC AAU17403;

DT 07-NOV-2001 (first entry)

DE Novel signal transduction pathway protein, Seq ID 968.

XX Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;
 KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
 KW immune system disorder; rheumatoid arthritis; inflammatory condition;
 KW organ transplant rejection; infection; hepatitis C; blood disorder;
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
 KW reproductive system; gastrointestinal; liver disorder; AIDS;
 KW acquired immune deficiency syndrome.

OS Homo sapiens.

PN WO200154733-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01312.

XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228927.
PR 01-SEP-2000; 2000US-0229284.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231444.
PR 08-SEP-2000; 2000US-0231444.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251038.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-465460/50.
N-PSDB: AAS27320.

Novel polypeptides useful for diagnosing, treating, preventing and/or
prognosing disorders related to the proteins, including cancers, immune
disorders and neuronal disorders -

Claim 1; SEQ ID No 968; 880bp; English.

The invention relates to novel isolated polypeptides (I), and
polynucleotides (II). (I), (II) and the antibody to (I) are useful for
diagnosing, preventing and treating diseases including immune system
disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
transplant rejections and graft versus host disease, infectious diseases
(e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
other blood-related disorders (sickle cell anaemia), myeloproliferative
disorders, primary haematopoietic disorders (leukemia), myeloproliferative
disorders (e.g. Gaucher's disease and cancer), neurodegenerative
disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal

CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 14.4%; Score 195; DB 22; Length 273;
Best Local Similarity 100.0%; Pred. No. 6.8e-194;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 CTVRIKDGLRNRFRQAOHMFROAMRTPIIMFHVPAANKEDYEOLSOSEKNYYSSRF 382
DB 65 CTVRIKDGLRNRFRQAOHMFROAMRTPIIMFHVPAANKEDYEOLSOSEKNYYSSRF 124
QY 383 SPDSQYIDNRSVNSAGLHTVQARPLNHPPEQIDSHSRPLPHSAHPGKPPSAPAPQNV 442
DB 125 SPDSQYIDNRSVNSAGLHTVQARPLNHPPEQIDSHSRPLPHSAHPGKPPSAPAPQNV 184
QY 443 FSTTVSSGVYFKKIKRNLQIKKGTGCLGFSITSRDVTIGGSAPYVKNILPRGAATD 502
DB 185 fstvssgytkkikrnlqikkgtegigfsitrdvtlgsaplyvknllprgaatgd 244
QY 503 GRLKAGDRIEYNGV 517
DB 245 grlkagdrilevngv 259

RESULT 5
ABG04716
ID ABG04716 standard; Protein; 419 AA.
XX
AC ABG04716;

DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #4707.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001MO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS68903.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 35075; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX

Sequence 419 AA;

Query Match 8.7%; Score 118; DB 22; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.5e-113;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1026 FGHRKDDKTEKTKIKIOESFTSEERIRMKOERIOAKTRFRROARERDYAEIOD 1085

DB 113 fgkhrkdkkiektykligesftseerirmkqeritqaktrfrqerayaeiqd 172

QY 1086 FHRTFGCDDIELMYGVSSEYEGSMALNARPOSPPRGHMDALYAQVKKPRNSKPSVDS 1143

DB 173 fhrtfgcdelmygvsyegsmalnarpqspregmhmdalyaqvkkprnskpsvds 230

RESULT 6
AAV71410
ID AAV71410 standard; Protein; 744 AA.
XX
AC AAV71410;

DT 04-OCT-2000 (first entry)

XX PDZ domain-comprising PHIP (ephAin Interacting Protein) complex.

XX PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW cellular process; nerve cell interaction; regeneration of nerve cell;
KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;
KW differential disorder; PHIP; ephrin interacting protein; syntenin.
XX

OS Mus sp.

XX Key Location/Qualifiers

XX Domain 273..351

XX Domain /label= PDZ_domain

XX Domain 461..546

XX Domain /label= PDZ_domain

XX Domain 590..678

XX Domain /label= PDZ_domain

XX WO200031124-A2.

XX 02-JUN-2000.

XX 19-NOV-1999; 99WO-CA01101.

XX 20-NOV-1998; 98US-0109158.

XX (MOUN) MOUNT SINAI HOSPITAL.

XX Lln D, Pawson A;

Thu Jul 25 08:38:16 2002

us-09-757-781-2.olg.rag

XX WPI: 2000-400038/34.
 DR N-PSDB: AAD01182.
 PT Isolated complex for treating proliferative or differentiative
 disorders comprises B class ephrin and PDZ domain containing protein -
 XX Claim 26: Fig 2D: 59pp: English.
 PS
 CC The patent discloses a complex comprising of a B class ephrin and PDZ
 CC domain containing protein. B class ephrins function as ligands for Eph
 CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
 CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
 CC binding site. This complex is used in methods to modulate the interaction
 CC of a B class ephrin and PDZ domain containing protein and to identify
 CC modulator compounds. It is also used for modulating cellular processes
 CC like, axonogenesis, nerve cell interactions and regeneration of nerve
 CC cells. The complex is also useful for treating proliferative or
 CC differentiative disorders associated with this protein complex.
 CC The present sequence is the PHIP (ephrin interacting protein) complex.
 CC comprising three PDZ domains. PHIP cDNA is isolated by screening the
 CC lambdaEx10x 10.5 day mouse embryo expression library. PHIP is closely
 CC related to Caenorhabditis elegans PAR-3 protein, involved in regulating
 CC polarity of the early embryo. PHIP sequence can be used as a probe, to
 CC isolate other proteins like GRIP and syntenin.

XX Sequence 744 AA:
 SQ

Query Match 8.1%; Score 110; DB 21: Length 744;
 Best Local Similarity 100.0%; Pred. No. 5.5e-105; Indels 0; Gaps 0;
 Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 39 AKDPYWIQVHRLEHSGGILDDILICVADDKRLVAVPDEQPHHGSGTSSNGT 98
 DB 39 akdpwywlvhrlehgddgllldilicvaddkrlvavpdeqphhgsgtssngt 98
 QY 99 GSPETIGSELGTNNVSAPQYQATSETEVPVYLRANMPLHVRSSDPAL 148
 DB 99 gspetigselgtcnvsatfpqatselevpsvlyranmplhvrssdpal 148

RESULT 7
 ABG04714 standard; Protein: 193 AA.
 ID ABG04714;
 AC ABG04714;
 XX 13-FEB-2002 (first entry)
 DT Novel human diagnostic protein #4705.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW Human; food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drimanac RT, Liu C, Tang YF.
 PI WPI: 2001-639362/73.
 DR N-PSDB: AAS68901.
 XX

PI New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20: SEQ ID No 35073; 103pp: English.

PS The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (II) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 193 AA:
 SQ

Query Match 5.9%; Score 80; DB 22: Length 193;
 Best Local Similarity 100.0%; Pred. No. 2.8e-74; Indels 0; Gaps 0;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1063 IOAKTRFERERQARERDYAEIOFHRTFCDELMYGVSSYEGSMALNRPSPBEGH 1122
 DB 57 iqaktrfereqaretyaeiofhrtfcdeelmvgvssyegsmalnarpspbeghm 116
 QY 1123 MDALYAGVKKPRNSKPSPPVD 1142
 DB 117 mdalyagvkkprnskpsppvd 136

RESULT 8
 ABB43115 standard; Peptide: 60 AA.
 ID ABB43115;
 AC ABB43115;
 XX 04-FEB-2002 (first entry)
 DT Peptide #10621 encoded by human foetal liver single exon probe.
 DE Human; foetal liver; gene expression; single exon nucleic acid probe.
 KW Homo sapiens.
 OS Homo sapiens.
 PN WO200157277-A2.
 PD 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US00669.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR.
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX
XX Claim 27; SEQ ID NO 35750; 639pp + sequence listing; English.
XX
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human fetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 60 AA;

Query Match 4.4%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 6.4e-54;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 LVAVFEDQDPHHGGDGTSSASTGTQSPETFGSELGTNNVSAFQPYQATSELEVTPSVLRA 134
Db 1 lvaifdeqdpnhggdgtssastgtqspetfgselgtnnvsafqpyqatselevtpsvlra 60

RESULT 9

AAAM36946
ID AAM36946 standard; Protein; 60 AA.
XX
AC AAM36946;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #10983 encoded by probe for measuring placental gene expression.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
XX Homo sapiens.
OS
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PT 30-JAN-2001; 2001MO-US00663.
XX
PF 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID NO 37215; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP;
CC see AAI31315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX
SQ Sequence 60 AA;

Query Match 4.4%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 6.4e-54;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 LVAVFEDQDPHHGGDGTSSASTGTQSPETFGSELGTNNVSAFQPYQATSELEVTPSVLRA 134
Db 1 lvaifdeqdpnhggdgtssastgtqspetfgselgtnnvsafqpyqatselevtpsvlra 60

RESULT 10

AAV71424
ID AAV71424 standard; Protein; 90 AA.
XX

AC AAV71424;
XX
DT 04-OCT-2000 (first entry)
XX

DE PDZ domain-comprising PHIP PDZ3 protein complex.

XX PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW cellular process; nerve cell interaction; regeneration of nerve cell;
KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;
KW differentiative disorder; PHIP PDZ3 complex; ephrin interacting protein.
XX

OS Mus sp.

PN WO200031124-A2.
XX

PD 02-JUN-2000.
XX

PF 19-NOV-1999; 99WO-CA01101.
XX

PR 20-NOV-1998; 98US-0109158.
XX

PA (MOUN) MOUNT SINAI HOSPITAL.
XX

PI Lin D, Pawson A;
XX

DR WPI; 2000-400038/34.
XX

PT Isolated complex for treating proliferative or differentiative
PT disorders comprises B class ephrin and PDZ domain containing protein -
XX

PS Claim 3; Fig 2C; 59pp; English.
XX

XX The patent discloses a complex comprising of a B class ephrin and PDZ
CC domain containing protein. B class ephrins function as ligands for Eph
CC domain tyrosine kinases (RTK) and possess a transmembrane element and a
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC binding site. This complex is used in methods to modulate the interaction
CC of a B class ephrin and PDZ domain containing protein and to identify
CC modulator compounds. It is also used for modulating cellular processes
CC like, axonogenesis, nerve cell interactions and regeneration of nerve
CC cells. The complex is also useful for treating proliferative or
CC differentiative disorders associated with this protein complex.
CC The present sequence is the PDZ domain-comprising PHIP (Ephrin
CC Interacting Protein) PDZ3 protein complex. PHIP consists of three PDZ
CC domains and is closely related to Caenorhabditis elegans PAR-3 protein,
CC involved in regulating polarity of the early embryo. This sequence
CC is isolated by screening the lambdaEx10x 10.5 day mouse embryo
CC expression library.
XX

SQ Sequence 90 AA;

Query Match 4.4%; Score 60; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 9,2e-54;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 594 NDSCGAGVGVKRNKRNENHADIGIVKSIINGGAASKDGRIRVNDQILAVNGESLIGK 653
Db 6 ndscgaglvavkvknrskenhadlglfivkslinggaaskdgrlrvndqilavngesllgk 65

RESULT 11

ABG04715
ID ABG04715 standard; Protein; 169 AA.

AC ABG04715;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #4706.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YF;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS68902.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT biologically for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 35074; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG030377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 169 AA;

Query Match 3.58%; Score 47; DB 22; Length 169;
Best Local Similarity 100.0%; Pred. No. 5,6e-40;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1144 NRSTPSNHDRIQRIQRFQFOAKODEVEDRRTYSFEQPPNARPAW 1190
Db 85 nrstpsnhdrlqrlrqfqqakodevedrtrrtystqppnarpaw 131

RESULT 12

AAG02888
ID AAG02888 standard; Protein; 51 AA.

AC AAG02888;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 6969.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR N-PSDB; AAC02894.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PT Claim 13; SEQ ID 6969; 71pp + CD-ROM; English.

CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC ends are can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.

Sequence 51 AA;

Query Match 2.8%; Score 38; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 4,4e-31;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1023 WFRFGRKHKDKIRKTKIKIQESFTSEERIRKQDQ 1060
Db 14 mfrfgkhrkdkikrktikikiqesftseerirmkqg 51

RESULT 13

AAV71423
ID AAV71423 standard; Protein; 86 AA.
XX
XX AAV71423;
XX
DT 04-OCT-2000 (first entry)
XX
DE PDZ domain-comprising PHIP PDZ2 protein complex.
XX
KW PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW cellular process; nerve cell interaction; regeneration of nerve cell;
KW axonogenesis; antiproliferative; proliferative disorder; treatment;
KW differentiative disorder; PHIP PDZ2 complex; ephrin interacting protein.
XX
OS Mus sp.
PN
XX W0200031124-A2.
XX
PD 02-JUN-2000.
XX
PE 19-NOV-1999; 99WO-CA01101.
PR 20-NOV-1998; 98US-0109158.
XX
PA (MOUN) MOUNT SINAI HOSPITAL.
XX
PI Lin D, Pawson A;
DR WPI; 2000-400038/34.
XX
PT Isolated complex for treating proliferative or differentiative
disorders comprises B class ephrin and PDZ domain containing protein -
PS Claim 3; Fig 2C; 59pp; English.
XX
XX The patent discloses a complex comprising of a B class ephrin and PDZ
domain containing protein. B class ephrins function as ligands for Eph
receptor tyrosine kinases (RTK) and possess a transmembrane element and a
highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
binding site. This complex is used in methods to modulate the interaction
of a B class ephrin and PDZ domain containing protein and to identify
modulator compounds. It is also used for modulating cellular processes
like, axonogenesis, nerve cell interactions and regeneration of nerve
cells. The complex is also useful for treating proliferative or
differentiative disorders associated with this protein complex.
XX The present sequence is the PDZ domain-comprising PHIP (Ephrin
interacting protein) PDZ2 protein complex. PHIP consists of three PDZ
domains and is closely related to Caenorhabditis elegans PAR-3 protein,
involved in regulating polarity of the early embryo. This sequence
is isolated by screening the lambdaEx10x 10.5 day mouse embryo
expression library.
XX
XX
XX Sequence 86 AA;
XX

```

Query Match      2.3%; Score 31; DB 21; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.4e-23;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 489 YKKNLIPFGAALIQDGRLAAGDRLILRYNGVDL 519
      |||||||
Db 29 yvknllprgaalqdgrlkagdrllleavngvdl 59

RESULT 14
ABG04627
ID ABG04627 standard; Protein: 187 AA.
XX
AC ABG04627;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4618.

```

XX	
KM	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-Oct-2001.
XX	
PE	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
XX	N-PSDB; MAA668814.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -

PS Claim 20; SEQ ID NO 34986; 103bp; Enfl15h.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct/sequences.

```

Query Match      0.8%; Score 11; DB 22; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 SELCTNNYSAP 116
      |||||
Db 36 selginnvsat 46

```

```

RESULT 15
AB111149
ID AB111149 standard; peptide: 315 AA.
XX
XX AC AB111149:
XX
DT 11-JAN-2002 (first entry)
XX
XX Human PAR3 homologue, SEQ ID NO:1519
XX
DE

```

Human; cytokine; cell proliferation; cell differentiation; growth factor;
 Haematopoiesis regulation; tissue growth; immunomodulator; activin;
 Inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 myeloid cell disorder; lymphoid cell disorder; aplasia; arthritis;
 chronic inflammatory condition; proliferative retinopathy;
 atherosclerosis; coronary heart disease; arterial ischaemia;
 bone disorder; osteoporosis; vascular growth disorder;
 tissue regeneration; wound healing; infection; immune disorder;
 cell culture; drug screening; gene therapy; antiinflammatory;
 antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 cyostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 antifungal; vulnereary; antitumor.

Homo sapiens.

WO200157188-A2.

09-AUG-2001.

05-FEB-2001; 2001WO-US03800.

03-FEB-2000; 2000US-0496914.

27-APR-2000; 2000US-0560875.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Dymnac RT;

WPI; 2001-457740/49.

N-PSDB; ABA08393.

Human proteins and DNA encoding sequences useful for preventing,
 treating or ameliorating a medical condition in a mammalian subject
 e.g. arthritis and cancer.

Claim 20; Page 150; 1963pp; English.

Sequences ABA10981-ABA12330 represent 1350 novel human polypeptides, and
 sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 invention also relates to vectors and recombinant host cells comprising a
 nucleotide of the invention, methods of producing the novel polypeptides,
 antibodies against the polypeptides, methods of detecting the nucleotides
 or polypeptides in a sample, and methods of identifying compounds which
 bind to polypeptides of the invention. Although novel, many of the
 polypeptides of the invention have homology to known proteins, thereby
 giving an insight into their probable biological activities, and hence
 potential therapeutic applications. The polypeptides of the invention may
 have various activities, including cytokine, cell proliferation or cell
 differentiation activities; stem cell growth factor activity;
 haematopoietic regulatory activity; tissue growth activity;
 immunomodulatory activity; activin- or inhibin-related activities;
 chemotactic or chemokinetic activities; haemostatic, thrombotic or
 thrombolytic activities; receptor or ligand activities; or may be
 involved in oncogenesis, cancer cell proliferation or metastasis.
 Depending on their biological activities, polypeptides and nucleotides of
 the invention are useful for preventing, treating or ameliorating medical
 conditions, e.g., by protein or gene therapy. Such conditions include
 cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 proliferative retinopathy, atherosclerosis, coronary heart disease,
 arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 vascular growth. Polypeptides involved with tissue regeneration and
 repair (or nucleic acids encoding them) may be used to promote wound
 healing (e.g., of burns, incisions and ulcers), while those with
 immunomodulatory activities may be used in the treatment of viral,
 bacterial and fungal infections in addition to immune disorders.
 Polypeptides with growth factor activity may be used in cell cultures to
 promote cell growth. For example, such polypeptides may be used to
 manipulate stem cells in culture to give rise to neuroepithelial cells
 that can be used to augment or replace cells damaged by illness,
 autoimmune disease or accidental damage. The polypeptides and nucleotides
 may also be used in the diagnosis of the above conditions, and in drug

CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention.

SQ Sequence 315 AA;

Query Match 0.8%; Score 11; DB 22; Length 315;

Best Local Similarity 100.0%; Pred. No. 0.03;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 GGPLGIHVVPF 291

DB 39 GGPLGIHVVPF 49

Search completed: July 24, 2002, 14:34:44
 Job time: 235 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 06:39:23 ; Search time 3972.89 Seconds
(without alignments)
19327.033 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689

Sequence: 1 atgaagtgacgtgtcctt.....tcttcctgtaaaaaaaaa

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapept 1.0

Searched: 13736207 seqs, 6748477542 residues

27472414

Total number of hits satisfying chosen parameters:

27472414

Minimum DB seq length: 0

27472414

Maximum DB seq length: 200000000

27472414

Post-processing: Minimum Match 0%

27472414

Maximum Match 100%

27472414

Listing first 45 summaries

27472414

Database : EST:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	936.6	16.5	1078	9	AL529631
C 2	886.2	15.6	1074	9	AL529631
C 3	765	13.4	870	10	BG743714
C 4	761.2	13.4	767	10	BG1917860
C 5	757.4	13.3	1150	10	BM466811
C 6	734.6	12.9	1063	10	BM462581
C 7	714.6	12.6	852	10	BG745599
C 8	710.8	12.5	725	10	BE792557
C 9	686.6	12.1	713	10	BG170486
C 10	665.8	11.7	1040	10	BG745723
C 11	665	11.7	973	9	AL529632
C 12	658.2	11.6	694	10	BG910244
C 13	644.6	11.3	760	10	AL529631
C 14	644	11.3	945	9	AL529635
C 15	638.2	11.2	966	10	BG260547
C 16	637.8	11.2	941	10	BG336977
C 17	626.4	11.0	687	9	BE176633

C 18	605.2	10.6	645	10	BE693704
C 19	603.4	10.6	779	10	BG209378
C 20	600.8	10.6	985	10	BG122324
C 21	599.2	10.5	987	10	BG170054
C 22	599	10.5	608	9	AM452651
C 23	594	10.4	617	10	BG431070
C 24	586.4	10.3	599	9	AM780417
C 25	579.8	10.2	618	9	BE176440
C 26	573.2	10.1	592	9	AL599694
C 27	573	10.1	660	10	BG473458
C 28	571.4	10.0	574	10	BG470230
C 29	557.8	9.8	814	9	AU079784
C 30	554.6	9.7	561	9	AU079061
C 31	553.4	9.7	718	10	BG206726
C 32	553	9.7	575	10	B1870563
C 33	552.6	9.7	626	10	BF679962
C 34	549.6	9.7	740	10	BG184956
C 35	538.4	9.5	699	9	BB545982
C 36	536.8	9.4	971	10	BG261114
C 37	536	9.4	536	9	AU079538
C 38	536	9.4	536	9	AU079538
C 39	536	9.4	809	10	BI093826
C 40	530.4	9.3	540	9	AM242451
C 41	526.2	9.2	545	9	AI129988
C 42	515	9.1	535	10	BE870124
C 43	511.4	9.0	534	9	AM835321
C 44	507.4	8.9	600	10	BE299145
C 45	504.6	8.9	1207	10	BF581891

ALIGNMENTS

RESULT 1
AL529631/c
LOCUS AL529631 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD0051D18 3
DEFINITION prime, mRNA sequence.
ACCESSION AL529631
VERSION AL529631.1 GI:12793124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1078)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
TITLE
COMMENT
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source
1..1078
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD0051D18"
/clone_1lb="LTI_NFL001_NBC4"
/sex="male"
/tissue="neuroblastoma cells"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com" 6 others
BASE COUNT 236 a 283 c 222 g 331 t

Qy 2844 ggaagaa 2850
|||||
Db 845 GGAAGAA 851

RESULT 8
BE792557 725 bp mRNA linear EST 20-SEP-2000
LOCUS 601585341p1 NIH_MGC_7 Homo sapiens CDNA clone IMAGE:3939370 5',
DEFINITION mRNA sequence.
ACCESSION BE792557
VERSION BE792557.1 GI:10213755
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Lung Hong/Rubln Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM788 row: b column: 11
High quality sequence stop: 725.
Location/Qualifiers
1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3939370"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 256 a 118 c 207 g 144 t

ORIGIN

Query Match 12.5% Score 710.8; DB 10; Length 725;
Best Local Similarity 99.6% Pred. No. 1.8e-114;
Matches 723; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 2597 tgaatgacacaaagcaggttcctccacagagatgtggtcttcctccctggtgctgaaga 2656
|||||
Db 1 TGGATGACACAAAGCAGGTTCTCCACAGAGATGTGGTCTTCCCTGGGTCTGAAGA 60

Qy 2657 agtcaagctcgttggagagcttgcagaccgagttgcgaggtgaacttgaatgggagaa 2716
|||||
Db 61 AGTCAAGCTCAATTGGAGAGTGTGACAGCGCAGTTGCCGAGGTGACTTTGAATGGGANA 120

Qy 2717 ttcccttcacatcgtccaagcgccggagataatcagaagggagggatgaatgaagacttca 2776
|||||
Db 121 TTCTTTCATCTCCATCGTCACGCGCGGATATCATGAGCGGATGCAATGAGACTTCA 180

Qy 2777 gagctgcacatgcacaactcttatgataaacccggtgtagatgatgatgaagcatg 2836
|||||
Db 181 GAGCTGCATGACAAATCTTATGTATTAACCCGCGGTAGATATATGATGAAGGATG 240

Qy 2837 agaccttgaagaagaacacagaagaagatcgaagatcaggagagagctglatccacg 2896
|||||
Db 241 AGACCTTGAAGAAGACACAGAAAGATTCAAGATCAGGAGAGAGCTGTATCCACAG 300

Qy 2897 ccaagtatcagcttccctcctctctggaagagacaattgaattggaaccagaagaaagc 2956
|||||
Db 301 CCAGTGATCAGCTTCCCTCCACTCTCTGAGAGACAAATGATGMAACCAAGAGAAAGGTG 360

Qy 2957 ataagactgatagaanaaggaataaactggaagaagaagaagaagaatagatgaag 3016
|||||
Db 361 ATTAAGACTATGAAAAAAGATTAACCTGAAAAAGAAAGAAAGATAGAGATAGG 420

Qy 3017 agaagataaataaagccaaaggaaggaatgctgaagggcttggagacatgltcaggt 3076
|||||
Db 421 AGAAGATTAATAATGAAGCCCAAGAAAGGAATGCTGAAGGGCTTGAGACATGTCAGGT 480

Qy 3077 ttggcacaacatcgaaagatgtgacaagattggaagaaacgggttaataataataacgaagt 3136
|||||
Db 481 TTGGCAACATCGAAAAAGATGACAAAGATTGAGAAAAAGGGTAAATTAATACAGGAAT 540

Qy 3137 ccttaccacagaagagagagatagatgaagcagagcagagagagattcaagcca 3196
|||||
Db 541 CCTTACATCAGAACAGAGAGAGATACGAATGACAGACAGACAGAGATTCAAGCCA 600

Qy 3197 aaactcgaaattagaggaacagcaagctcgagagcgltgactgtgaattcaagatt 3256
|||||
Db 601 AAACTCGAATTTTAGGCAAGCAAGCTCGAAGCGTCACTATGCTGAATCAAGATT 660

Qy 3257 ttcatcgacatttgctgtgatatgaattaatgtaatggggagcttcttataagaag 3316
|||||
Db 661 TTCATCGCACTTTGG-TGTGATGATGAGATTATGTATGGGGAGTCTTCTTATGAAG 719

Qy 3317 gtccca 3322
|||||
Db 720 GTTCCA 725

RESULT 9
BG170486 713 bp mRNA linear EST 06-FEB-2001
LOCUS BG170486
DEFINITION 602322859p1 NIH_MGC_89 Homo sapiens CDNA clone IMAGE:4425975 5',
LOCUS mRNA sequence.
ACCESSION BG170486
VERSION BG170486.1 GI:12677189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 713)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10172 row: e column: 16
High quality sequence stop: 710.
Location/Qualifiers
1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4425975"
/clone_lib="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;

Page 8

	NCBI Value	Mismatches	Indels	Gaps
DY	3099	0;	4;	1;
	caagattgcgaacaggtaaatataaacaggaatccttatacagaacc-			
Dδ	1	CACATTGGAAGAACCGTCTTTGGATCCTTGTCCCTTCTGTTTC		

Page 8

FEATURES

BASE COUNTY
ORIGIN

Matches

111

Db 579 ---GTTCTCCAGCAGAGATGTGGTCTCTCCCTGGTCTGTAAGAATGTCAMGTCATTGG 523
Qy 2672 agagctcagacacgcaagctgcagagctgacttgaaatgaggaattcttcacatc 2731
Db 522 AGAGCTCTCAGACCCGCACTTCCGAGGTGACTTTCAATGGGATATTTCTTCCATCGTC 463
Qy 2732 caagcgcgagataatcagaagcaagggatcaatgaaagcttcagaagctgcacagaa 2791
Db 462 CACGCCCCGGATATATCAGAGGACAGGATCAATGAGAGCTTCAGACCTGCCATCGACA 403
Qy 2792 aatcttaataaaccggttagatgatatgataaggaatgaggaacgttgagaaag 2851
Db 402 AATCTTAATGATTAACCCCGGAGATGATGATGATGATGATGATGATGATGATGATGAT 343
Qy 2852 acacagaagaagttcaagatcagagagagagctgtatccacagcagctgacagcct 2911
Db 342 ACACAGAAAGATTCAGAGATCAGGAGAGAGTGTGTATCCACACAGCTGATCAGCCTT 283
Qy 2912 cccactctgagagagacaaatgaaatgaaacaaagaaagtgataagactgataaa 2971
Db 282 CCCACTCTCTGAGAGCAAAATGAAATGCAAAACCAAGAGAAAGTGTAAAGCTGATAGAA 223
Qy 2972 aaaaagataaacctgagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 3031
Db 222 AAAAGGATTAAGCTGGAAGAAAGAAAGAAAGATGATGATGATGATGATGATGATGATGAT 163
Qy 3032 aagcaagaaggaatgctgaagagcttgagagacatgttcaggttgcaacaatcga 3091
Db 162 AAGCAAGAAAGGAATGCTGAAGGCTTGGGAGACATGTTCAGCCTTGCCAAACTGAAG 103
Qy 3092 aagatgacaagattgagaaacgggtataataaat 3128
Db 102 CCCGAGAGAGATGAAACAAAGCAAGATTCAAACAT 66

RESULT 11
LOCUS AL529632 973 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529632 LTI_NFL001_NBC4 Homo sapiens CDNA clone CSDD005YD18 5
AL529632
ACCESSION AL529632
VERSION AL529632.1 GI:12793125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 973)
AUTHORS L.J.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length CDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..973
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD005YD18"
/clone_id="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand CDNA
was primed with a NotI-oligo(dt) primer. Five prime end
enriched, double-stranded CDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the PCMVSPORT 6
vector. Library was normalized. Library was constructed
by life technologies. Contact : Feng Liang life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifestech.com URL :

BASE COUNT 209 a 290 c 286 g 171 t 17 others
ORIGIN

Query Match 11.7%; Score 665; DB 9; Length 973;
Best Local Similarity 97.3%; Pred. No. 1.9e-106;
Matches 674; Conservative 11; Mismatches 7; Indels 1; Gaps 1;

Qy 1 atgaagatgacggctgctgctgagcagggagggctgctgctgagggagggc 60
Db 281 ATGAAGATTACCGTATKCTTTGGACGACCGGGGTGTGTCGGCGGAGCGCCAC 340
Qy 61 atgaagatttctagcctcaccagcagcggtgacccgctacaggaagccatcgcaag 120
Db 341 ATGAAAKTTTTCAGCCTCATCCAGCAGCGGGGTGACCCCGTACCGGAAGCCATCCGCAAG 400
Qy 121 gatcaaacactgatatagatgcatgcttggaacatgagatgaggaataactagac 180
Db 401 GATCCAAACTACTGATGATACAGGTGCTCCCTTGAAACATGAGATGAGGAATACTAGAC 460
Qy 181 ctgtatgacatctctgtgtatgtagagacgaataagacagactgttagatgtttat 240
Db 461 CTGTGATGACATTTCTTTGTATGTATGATGATGATGATGATGATGATGATGATGAT 520
Qy 241 gacagagatccacatcaccagagagtgatgacacagctgcccagttccacgggtaccagagc 300
Db 521 GAGCAGGATCCACATCAGCAGGATGATGATGATGATGATGATGATGATGATGATGAT 580
Qy 301 ccagagataattgtgtatgtagcttgagcaccacaatgcttcagccttacccttacc 360
Db 581 CCAGAGATATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 640
Qy 361 gcaacaagatgaaatgagatgacacacacacacacacacacacacacacacacacacac 420
Db 641 GCAACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 700
Qy 421 cgaagcagtagtgaac 480
Db 701 CGACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 760
Qy 481 tctcttgaagagccttcaagagaataatcccaacagctgtgtcaacaacagctgtctc 540
Db 761 TCCTCTGAAGACCTTCAAGGAAATATCCACACGCTGTGTCAACACACCTGCTTCC 820
Qy 541 aagcagaacacgct-gggatgacctaataacacacacacacacacacacacacacacac 599
Db 821 AAGCAAGACACTGCTGGGAGTCTTAACCTGCGACAGGAAGATGAAACTACAG 880
Qy 600 aagctcccgagagatagtaactgcttaacacacacacacacacacacacacacacacac 659
Db 881 AAGCCTCCCGGAGATACATGATGATGATGATGATGATGATGATGATGATGATGAT 940
Qy 660 gtctctgagtgccagtcaccacaatggtggca 692
Db 941 GTCTCTGATGCTGACGATCACCACCAATGCTGGGCA 973

RESULT 12
Bg910244
LOCUS Bg910244 694 bp mRNA linear EST 05-JUN-2001
DEFINITION 602805735P1 NCI_CGAP_Brn67 Homo sapiens CDNA clone IMAGE:4937865
5', mRNA sequence.
ACCESSION Bg910244
VERSION Bg910244.1 GI:14290720
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 694)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

us-09-757-781-20.rst

Unpublished (1999)
Contact: Robert Stammers

High quality sequence stop: 694.
Location/Qualifiers
1..694

186 a /note="Organ: brain; pi3k-resistance")
187 Site_2: SalI; Vector: pCMV-Sport6; Site_1: NotI;
188 Average insert size 2.3 kb. Constructed by Life
189 Technologies. Note: this is a NCI-CCAP Library."
205 g 114 t

Similarity	11.68;	Score	658.2;	DB	10;	Length	694;
Conservative	98.68;	Pred.	No.	3e-105;			

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

RESULT 1
BG195361

LOCUS	BG195361	760 bp	mRNA	1 linear	EST 21-APR-2001
DEFINITION	RSF14543	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.			
ACCESSION	BG195361				
VERSION	BG195361.1	GI:13717048			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (Phases 1 to 760)	Harrington, J. J., Sherif, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Costanzo, D., McGilgott, K., Boozar, S., Mays, R., Smith, J., Veloso, N., Kitz, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducart, M.	Creation of genome-wide protein expression libraries using random activation of gene expression	Nat. Biotechnol.	19 (5), 440-445 (2001)
2 (Phases 761 to 1520)	Harrington, J. J.	Primates; Catarrhini; Homidae; Homo.		
3 (Phases 1521 to 2280)	Harrington, J. J.	Euarchonta; Eulestomina;		

FEATURES
Source

```

/branch="Homo sapiens"
/closet1="taxon:9606"
/closet1b="Athersys RAGE Library"
/cell_line="Htt1080"
/Note="See 'Creation of Genome-wide Protein Expression
        Libraries Using Random Activation of Gene Expression,'
        Nature Biotechnology, in press. Note that even though the
        cell type indicated is Htt1080, since a random activation
        method was used, these sequence tags are not necessarily
        expressed in Htt1080 under normal circumstances."
BASE COUNT      258 a      139 c      197 g      164 t
RIRIN

```

Query Match
Best Result

	Local Similarity	97.9%	Score 64.6;	DB 10;	Length 760;
Matches 653;	Conservative	0;	Mismatches 14;	Indels	0;
				Gaps	
Qy	2419	tctttagtcacagatgttgaatccaaattcttgcatttcaacagagaagatttgaagtcag	2478		
Db	7	tcttttaacagtcacagctgtttgttccaaagcttctgttttcaacgaacagattttcagctcag	66		
Qy	2479	agtaigtacagaaaacgcacaaagcaatttcacagatgcagatgaatttgatttcgttaa	2538		
Db	67	agtaatgtcagaaaacgcccaaacgcatttttcacatcccgatcaatttgatttcgttaa	126		
Qy	2539	acacgaatatcaaaagcagtgatttagtatacgcgcagagacacaaactaaataacagtg	2598		
Db	127	acacgaataatcaaaaagcattgatattttttagatagctcagacgaagactaaataacagtg	186		
Qy	2599	gatgaccagaagaacaggtttccccaagaagatgttgagttccctccctggagtcgaagaag	2658		
Db	187	gatgactacagaagaacaggtttccccaagaagatgtagtggtctccctccggtatgaagaag	246		

2659 tcaagctcgttgagagatctgcagacgcagcttcgcagagtgacttgatcagtgagatact 2718
|||||
Db 247 TCAAGCTCGTGGTGAGCTGACAGCCGACGTGGCGAGTGACTTTCGAATGGGGATATT 306
2719 ccttccatcgtccacgcgcgcgataatcagagcagggatgcaatgagaagcttcaga 2778
|||||
Db 307 CCTTCCGTCGCCCGCCGCGGATATCAGAGCGAGGATGCAATGACAGCTTCAGA 366
2779 gctgcacatcgaacaattctatgatataaacccgcggtatgataatgatagaagcagag 2838
|||||
Db 367 GCTGCATCGACAAATCTTATGATTAACCCCGGTAGATGATGATGAAGGATGAG 426
2839 accttgaagaagacacagaaagttcaagatcagagagagagctgctatccacagcc 2898
|||||
Db 427 ACCTTGGAGAGACACAGAGAAAGTTCAAGATCAGAGAGAGATGCTGTATCCACACC 486
2899 agtatcagccttcccatctctgagagagacaatgaaatgaaaccaaagagaagatgat 2958
|||||
Db 487 AGTATTCAGCCTTCCCTCTCTGAGAGACAAATGAAAGCAACCAAGAGAAAGGTAT 546
2959 aagactgataagaaaaagataaaacttgaagaaagaaaaagaaagatatagataagag 3018
|||||
Db 547 AAGACTGATAGAAAAAAGATTAACCTGAAAAAGAAAAAGAAAGATAGATAGAGAG 606
3019 aagataaataagaaagcagaaagaggaatgctgaaaggcttggagacatgttcagttt 3078
|||||
Db 607 AAGGATAAATGAAAGCCAGAAAGGGAATGCTGAAGGGCTTGGAGACAAATTCAGCCT 666
3079 ggcaaac 3085
|||||
Db 667 GCCAAG 673

RESULT 14
AL529865 945 bp mRNA linear EST_13-FEB-2001
LOCUS AL529865 LTI_NFI001_NBC4 Homo sapiens cDNA clone CSDD005YH02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL529865
VERSION AL529865.1 GI:12793358
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 945)
AUTHORS Li, W. B., Gruber, C., Jessup, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
SOURCE
1..945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD005YH02"
/clone_lib="LTI_NFI001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-Oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 204 a 284 c 286 g 164 t 7 others

ORIGIN
Query Match 11.3%; Score 644; DB 9; Length 945;
Best Local Similarity 98.5%; Pred. No. 8.8e-103;
Matches 656; Conservative 4; Mismatches 5; Indels 1; Gaps 1;
1 atgaagtgaccgtgtgtcttcggaacgacccgggtgtgtcgtgtcgtgagcagccac 60
Db 280 ATGAAAGTADACCGTGTGCTTGCGAGCGACCCGGGTGTGTGTCCTGCGGGAGCGCCAC 339
61 atgaagtttcagcctcaatccagcagcggtgtgacccgtctacggaaaggtccacag 120
Db 340 ATGAAAGTTCACGCTCATCCAGCAGCGGGTGTGACCCCGTACCGGAAGCCATCCCAAG 399
121 gatccaactacttgatatacaggtgatcagcttggagacatgagatgagaaatactagac 180
Db 400 GATCCAAACTACTGATATACAGGTGCATGCTTGGAACATGAGATGAGAAATACTAGAC 459
181 ctgtatgacatcttctgtatgtatgatacagataaagacagactgtgtacagtgatgtat 240
Db 460 CTGTATGACATCTTGTGTGATGTACAGACGATTAABACAGACTGTAGCACTGTTGAT 519
241 gagcagatcacaatcaacgaggtgtatgtgacccagttgccagttccacgggtaccagagc 300
Db 520 GAGCAGATTCACATCAGCAGAGAGTGTGACACAGTGCACAGTCCACCGGTACCCAGAGC 579
301 ccagagataattgtgtatgtatgtgtgacccaacaatgtctcagccttcagccttaccaa 360
Db 580 CCAGAGATATTTGGTAGTGACGCTTGCCACCAACAAATGTTCAGCCTTTGACGCTTACCA 639
361 gcaacaagtgaatgtgaagttcacaccttcagtccttcggaacaatagtctcttcagt 420
Db 640 GCAACAAGTGAATGTAGAGTTCACACCTTCAGTCTTGCAAAATATGCTTTCATGTT 699
421 cgaagcagtagtgagaccagcttaattgtgccttcacacttcgtatgtatgtatgtt 480
Db 700 CGACCATATAGTGTGACCCACACCTCTAATGTGCTCTCCTCCTCATCTGTGATATGTAATTT 759
481 tcccttgaagagccttcaagagaaaaatccacagcgtgtgtaacaacaagctgtctctc 540
Db 760 TCTCTGAAGAGACCTTCAAGGAAAAATCCACACCTGTGTCAACAACAGCTGGCTTCTC 819
541 aagcagaacactgtct-ggtagtctctaanaacctgcagcagaaagaaatgaataactacag 599
Db 820 AAGCAACACTGTCTBGGGAGTCTTAACCTGCGACAGAGAAAGATGAATAACTACAG 879
600 aagctccgcggggatctagttaactgtgtcttaaccaattccagagagacatgtcgtc 659
Db 880 AAGCTCCCGCGGATACTAGTAAGTGTCTAACAATTTTCAGAGAGAMAATGCTCGCTC 939
660 gtctct 665
|||||
Db 940 GTCTCT 945

RESULT 15
BG260547 966 bp mRNA linear EST_13-FEB-2001
LOCUS BG260547 602372096P1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4480366 5,
DEFINITION mRNA sequence.
ACCESSION BG260547
VERSION BG260547.1 GI:12770363
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 966)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10313 row: 0 column: 23
 High quality sequence stop: 723.
 Location/Qualifiers

FEATURES

source

1.966
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4480366"
 /clone_1lb="NIH_MGC_93"
 /issue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."

BASE COUNT 282 a 209 c 271 g 204 t
 ORIGIN

Query Match 11.2%; Score 638.2; DB 10; Length 966;
 Best Local Similarity 91.0%; Pred. No. 9,1e-102;

Matches 762; Conservative 0; Mismatches 53; Indels 22; Gaps 7;

OY 1774 ccacttaatgatcaggatctgcaagccttggtgctcagtgcaaggtcaacggtcaaaa 1833
 |||||||
 Db 1 CCACCTAATGATTCAGGATCTGAGGCTTGCTGTCAGTCAAGGTAAAGSTAAACGGTCAAAA 60
 OY 1834 ggaagaccgagatttggaatctttgcaagtcattatataatgaggaagcaatct 1893
 |||||||
 Db 61 GAGAACCCAGCATTTGGGAATCTTTGCAAGTCCATTATTAATGAGAGCAGCATCT 120
 OY 1894 aaagatggaaggtctcggtgaatgatacaactgataagcaatgaggaatccctgttg 1953
 |||||||
 Db 121 AAAGATGGAAGGCTGCGGTGCAATGATCAACGTATAGCAGTAAATGAGAGATCCCTGTTG 180
 OY 1954 ggcagaagcaaaccaagatgcatgtaaacccaagaaggtctatgctactgaagcaat 2013
 |||||||
 Db 181 GGCAGACAAACCAAGTGCATGGAACCCCTAAGAAAGGTATGTCTACTGAGGCAAT 240
 OY 2014 aaacgaggaatgatacagcttatgttgcaaggagaataagcaagtgcattgagctgaag 2073
 |||||||
 Db 241 AAACGAGGATGATCCAGCTTATTTGTCAGAGAGATAAGCAAGTCAATGAGCTGAAG 300
 OY 2074 tcaactggagcccccttgagacctgagctgcccattgaacaacagcgttgatgataagaa 2133
 |||||||
 Db 301 TCACCTGGAGACCCCTGGACCTGAGCTGCCATTGAAACGCGTTGGATGATAGAGAA 360
 OY 2134 cgaagaattccattccctctacagtggaatggaagggtctgataatgcgccacaga 2193
 |||||||
 Db 361 CGAAGAAATTTCCATTCCTCTACAGTGGATGAGGGGCTTGATGAATGCCACAGAGA 420
 OY 2194 aatgctgcctcagtaggataatggtgagtcagtaataacacagctgcccctaagctg 2253
 |||||||
 Db 421 AATGCTGCCCTCAGTAGATTAATG-----GGTAAATACAGCTGCTCCCTACAGTG 471
 OY 2254 aatatgcccaagaatgacatgcatatagaagaataacaggttgccagtgcttctcca 2313
 |||||||
 Db 472 AATTATGCCCAAGATGACATGTCATTATAGAGATGACAGGTTGCCAGTGCCTTCCCA 531
 OY 2314 catctctgaccagctctcttccaagctcccatatgatggtgggttggtgaagcagat 2373
 |||||||
 Db 532 CATCTCTCTGACAGTCTCTTCCAGCTCCATGATGATGAGG-TTGTACGCGAGAT 590
 OY 2374 gctgtactggcgaaggtcgcaatcagtgattcagccgactgctcttgatcagat 2433
 |||||||

Db 591 CCTGCTACTTGGGCCAA-GCTGCATCAGGTGATTCAGCCGACTGCTTTGAGTCCAGAT 649
 OY 2434 gtgatacag--ttcttgctttcaagaagaagatttgagcgtcaggtatgtaag--- 2488
 |||||||
 Db 650 GTAGATCCAGGCTTCTTGCTTTCAACGAAGAGGATTCGCCACGTCAAACTATGATGTC 709
 OY 2489 ---aaaacgacaaag--caatttcagatgccaagcaattgatttggttaaacag 2543
 |||||||
 Db 710 AGAAAAACGCACAAAGGCAAGTTTCAGATGCCAGTCAATGG-GTTCGCTAAACACGA 768
 OY 2544 aaatcaaaaagcattgattaggtatagctgacgaagactaaactcaataacagtgga 2600
 |||||||
 Db 769 AAGTTCAAAAGCTTGATTTAAGGTATAGCTGGAAGACTTAATCCAAAACGGGCTA 825

Search completed: July 24, 2002, 09:16:29
 Job time: 9426 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 06:47:48 ; Search time 101.24 Seconds
(without alignments) 13802.934 Million cell updates/sec

Title: US-09-757-781-20

Perfect score: 5689
Sequence: 1 atgaaagtacccgtgtcctt.....tcttcctgttaaaaaaaaaa 5689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PCrus.COMB.seq: *
6: /cgn2_6/ptodata/2/1na/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	2.2	1042	4 US-09-276-531-77	Sequence 77, Appl
2	105.6	1.9	7218	4 US-08-232-463-14	Sequence 14, Appl
3	58.8	1.0	7431	4 US-09-306-998-2	Sequence 2, Appl
4	52.2	0.9	8040	1 US-08-336-281-1	Sequence 1, Appl
5	52.2	0.9	8040	3 US-09-100-804-1	Sequence 1, Appl
6	51.4	0.9	87350	3 US-08-781-891-79	Sequence 79, Appl
7	51	0.9	6964	4 US-09-461-697-193	Sequence 193, App
8	51	0.9	6994	4 US-09-461-697-191	Sequence 191, App
9	51	0.9	7174	4 US-09-461-697-189	Sequence 189, App
10	51	0.9	7744	4 US-09-461-697-187	Sequence 187, App
11	51	0.9	8194	4 US-09-461-697-185	Sequence 185, App
12	51	0.9	16694	4 US-09-461-697-184	Sequence 184, App
13	50.6	0.9	8043	5 PCT-US94-09943-1	Sequence 1, Appl
14	50.6	0.9	8119	4 US-09-290-640-45	Sequence 45, Appl
15	49	0.9	5836	4 US-09-233-086-2	Sequence 2, Appl
16	47.8	0.8	11298	3 US-08-948-705-3	Sequence 3, Appl
17	47.8	0.8	11721	4 US-09-026-039-3	Sequence 3, Appl
18	47.8	0.8	11721	4 US-09-026-039-3	Sequence 3, Appl
19	47.2	0.8	3762	2 US-08-623-906A-18	Sequence 18, Appl
20	46.8	0.8	2447	2 US-09-014-969-14	Sequence 14, Appl
21	46	0.8	5552	3 US-08-155-888-1	Sequence 1, Appl
22	45.8	0.8	176373	3 US-09-128-155-17	Sequence 17, Appl
23	45.6	0.8	712	1 US-08-410-804-4	Sequence 4, Appl
24	45.6	0.8	712	1 US-08-259-514-4	Sequence 4, Appl
25	45.6	0.8	712	2 US-08-858-311-4	Sequence 4, Appl
26	45.6	0.8	1830	1 US-08-410-804-2	Sequence 1, Appl
27	45.6	0.8	1830	1 US-08-259-514-2	Sequence 2, Appl

28	45.6	0.8	1830	2 US-08-858-311-2	Sequence 2, Appl
29	45.6	0.8	15144	3 US-08-458-434A-6	Sequence 6, Appl
30	45.6	0.8	73604	4 US-09-268-992-7	Sequence 7, Appl
31	45.2	0.8	3064	4 US-09-461-697-203	Sequence 203, App
32	45.2	0.8	152331	3 US-09-128-155-16	Sequence 16, Appl
33	45	0.8	2223	1 US-08-257-073-4	Sequence 4, Appl
34	44.8	0.8	240	1 US-08-628-417-6	Sequence 6, Appl
35	44.6	0.8	2894	4 US-09-007-005-17	Sequence 17, Appl
36	44.6	0.8	2894	4 US-09-244-796-17	Sequence 17, Appl
37	44.6	0.8	3724	4 US-09-018-584A-13	Sequence 13, Appl
38	44.6	0.8	6755	3 US-08-931-999-4	Sequence 4, Appl
39	44.4	0.8	5433	3 US-08-929-329-1	Sequence 1, Appl
40	44	0.8	494	2 US-08-332-768A-22	Sequence 22, Appl
41	43.4	0.8	10514	4 US-09-245-041-10	Sequence 10, Appl
42	43.4	0.8	12047	2 US-09-022-461-1	Sequence 1, Appl
43	43.2	0.8	8920	4 US-08-446-855A-1	Sequence 1, Appl
44	43.2	0.8	8920	4 US-09-150-741-1	Sequence 1, Appl
45	43.2	0.8	152331	3 US-09-128-155-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-276-531-77
Sequence 77, Application US/09276531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESSES:
ADDRESSER: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,531
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/079,677
FILING DATE: March 27, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lynn E. Murry, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 1042 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BLADTUT04

```
; CLONE: 1555118
US-09-276-531-77

Query Match
Best Local Similarity 66.7%; Score 126; DB 4; Length 1042;
Matches 212; Conservative 0; Mismatches 100; Indels 6; Gaps 2;

QY 2510 cagatgcacatcattgattcgtttaaacacgaacaaatcaaaagatgattagta 2569
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 531 CACGGCGAGGACCCGTGATTCATTAATTGAAAAGCCGGAAGAGATGACCT---TG 587
QY 2570 tagctgcagagactaaactaatcagtgatgagccagaagccagctccacagag 2629
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 588 TGCCAGATGAAGCAAGCTTACATTCATGCTGACAAAATCGATCTCCAGCAAG 647
QY 2630 atgtgagctccctccgtctgaagaagtcgaagctgttgagagatctgcagccag 2669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 648 ATTTGGCTCCAACTGTGGTTTGAAGAAAGTCACGCTCTTGAGAGAGTGTGACAGCTG 707
QY 2690 ttccagagtgacttgaatgaggatattccttcacatgctgcagccgaggaatca 2749
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 708 TGCGCGAGGTCAGAGAGAAAT---GACCTTCCTTCAAGGCCCGCCGACATGGTTC 764
QY 2750 gaagcagaggatgcaatgagagcttcagagctgcacatgcacaaatctatgataaccg 2809
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 765 GAGGCCAGAGCTGCAATGAGAGCTTTAGAGAGCCATTGCAAAATCTAGATGACCTG 824
QY 2810 cagtagatgatgatg 2827
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 825 AAGAAATAGAGCTGACG 842

RESULT 2
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZpT-F15
US-08-232-463-14

Query Match
Best Local Similarity 5.1%; Score 105.6; DB 1; Length 7218;
Matches 21; Conservative 264; Mismatches 123; Indels 0; Gaps 0;

QY 2825 atgaagcattgagacttgaagaagacagcagaagaattcaagctggagagagt 2884
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1465 AAGTAGTTAAGATGAAGAAATTTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRR 1406
QY 2885 ctgatacagcagcagatgacgtctccactctctgcgagagacaatgaaacc 2944
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1405 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1346
QY 2945 aagagaaagtgataagactgatagaataaagataaactgagaagaagaagaag 3004
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1345 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1286
QY 3005 atagagataagagaagaagataaagcagaagaaggaatctgaaggcttggag 3064
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1285 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1226
QY 3065 acatgtcaggttgcacaacatcgaaagatgacgaatgagaacggtataataa 3124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1225 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1166
QY 3125 aaatacaggaatccttacatcagaagagagagatagaatgaacagagagagaga 3184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1165 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1106
QY 3185 ggaltaacccaacactcgaaatttagggaacgacagctcgagac 3232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1105 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1058

RESULT 3
US-09-306-998-2
; Sequence 2, Application US/09306998
; Patent No. 6291173
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Tavligian, Sean V.
; TITLE OF INVENTION: MMSC2- An MMAC1 Interacting Protein
; FILE REFERENCE: MMSC2
; CURRENT APPLICATION NUMBER: US/09/306,998
; CURRENT FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 60/084,740
; EARLIER FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 7431
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (57)..(6167)
US-09-306-998-2

Query Match
Best Local Similarity 1.0%; Score 58.8; DB 4; Length 7431;
Matches 129; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 1383 tatcagcttaagaaggtacagaaggttgggattcagcatcactccagagatgtaac 1442
```

Db 5918 tattaacactagcgagagcagatgcttagcttcagctatgctgagatagcgag 5977
Qy 1443 aatagtgctcagctcgaatcatalgtgaaaaacatctcccccggggggccgcatlca 1502
Db 5978 ccccatgagactaccatcattatgttaaaacagtgcttgaaaggagcagcctcga 6037
Qy 1503 ggaatggcgaacttaaggagagacagactatagaggttaagttagttagtgagg 1562
Db 6038 agacgagcgtctgaaaaaggcgatcagatcattgtctcattagggcagagctagaag 6097
Qy 1563 caatcccaagaagaagtgttgctgtgtgaaagcaccagaatgagaagactgtgag 1622
Db 6098 agtcaaccatgaagaagctgtgccaatcccttaaacgacaaaggcactgtcattgat 6157
Qy 1623 ccttct 1628
Db 6158 ggttct 6163

RESULT 4
US-08-596-291-1
: Sequence 1, Application US/08596291
: Patent No. 5821075

GENERAL INFORMATION:
APPLICANT: GONEZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,291
FILING DATE: 09-AUG-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: 10461/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIEL

SEQUENCE CHARACTERISTICS:
INFORMATION FOR SEQ ID NO: 1:
LENGTH: 8040 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HOMO SAPIENS
FEATURE:
NAME/KEY: CDS

LOCATION: 78..7475
US-08-596-291-1

Query Match 0.9%; Score 52.2; DB 1; Length 8040;
Best Local Similarity 51.5%; Pred. No. 0.00042;
Matches 120; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 1369 ggcagaagcttaatatccagcttaagaaggtaagaaggtttggatlcagcact 1428
Db 4107 GGAGATATCTTTGAGGCTTGACCTGCTAAATAATGATACAGCTTGCGGATGATGACG 4166
Qy 1429 tccagatgatacaatagtgctcagctccaatctatgtgaaaaacatctcccccgg 1488
Db 4167 GGAGGTGTGAATACAGTGTCAACATGCTGGCATTATGTATGTAAGATGTATTCGCCAG 4226
Qy 1489 gggcgccatccagatgcccacttaaggcagagagacagactataagtgaaatga 1548
Db 4227 GGAGCAGCAGACTCTGATGTGTGAATTCACAAAGGATGATGCTCTACTGCTCAATGGA 4286
Qy 1549 gtaattagtggaatcccaagaagaagtgttcgctgttgagaagcac 1601
Db 4287 GTTACTGTAGAAAGAGCCACCATAAGCAAGCTGTGTGAACACTGGAATATAC 4339

RESULT 5
US-09-100-804-1
: Sequence 1, Application US/09100804
: Patent No. 6066472

GENERAL INFORMATION:
APPLICANT: GONEZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: 10461/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8040 base pairs

TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: HOMO SAPIENS
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 78..7475
 US-09-100-804-1

Query Match 0.9%; Score 52.2; DB 3; Length 8040;
 Best Local Similarity 51.5%; Pred. No. 0.00042;
 Matches 120; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 1369 gccaaggcctaataatccagcctaagaaggtacagaaggttgaggttcagcact 1428
 Db 4107 GGAGTATCTTTGAGGTTGAAGTGGCTAAATATATAACACCTTGGGGATTAAGTCAAG 4166
 QY 1429 tccagagtgtaacataagtgaggtcagctcgaatctatgtgaaaaacattccccc 1488
 Db 4167 GGAGGTGTAATACAGAGTGTCAAGACATGTGGCATTATGTGAAGAAGTATTCGCCAG 4226
 QY 1489 gggggggccatcagatggcgcgacttaaggcagagagacagacttatagagtaatgga 1548
 Db 4227 GGAGCAGCAGAGTCTGAGTGAATTCACAAGGTGATCCGCTCCGTCGTCACATGGA 4286
 QY 1549 gtagattagtgaggcaaatcccaaggagaagttgttcgcgttgagaagcagc 1601
 Db 4287 GTTACTTAGAAGAGCCACCCATTAACCAAGCTGTGGAACACCTGAGAAATAC 4339

RESULT 6

US-08-781-891-79/c
 Sequence 79, Application US/08781891

Patent No. 6090620
 GENERAL INFORMATION:
 APPLICANT: Yu, Ying-Hui
 APPLICANT: Fu, Chang-Eu
 APPLICANT: Oshima, Junko
 APPLICANT: Mulligan, John T.
 APPLICANT: Schellenberg, Gerald D.
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 NUMBER OF SEQUENCES: 209
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/781,891
 FILING DATE: 27-DEC-1996
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6090620Ienbourg Ph.D., Carol
 REGISTRATION NUMBER: 39,317
 REFERENCE/DOCKET NUMBER: 240052.419
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 79:
 SEQUENCE CHARACTERISTICS:

LENGTH: 87350 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-781-891-79

Query Match 0.9%; Score 51.4; DB 3; Length 87350;
 Best Local Similarity 47.6%; Pred. No. 0.0036;
 Matches 151; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 2920 ctgagagacaaatgaatgaaacccaagaagaaggtgtaagactgtatagaataaagat 2979
 Db 55399 CTTGAAGAGAAAG 55340
 QY 2980 aaaaactggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 3039
 Db 55339 AAG 55280
 QY 3040 aagggaatgtaagggcttgaggacatgltcaggttggcaaacatcgaaaagatgac 3099
 Db 55279 AAGCGAGCGCGGTCCGACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55220
 QY 3100 aagattgagaacacgggttaataataacacggaatccttaccatcgaagaagagag 3159
 Db 55219 GAG 55160
 QY 3160 atagcaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3219
 Db 55159 GAG 55100
 QY 3220 caagctcgaagcgtga 3236
 Db 55099 AAGCGAG 55083

RESULT 7

US-09-461-697-193
 Sequence 193, Application US/09461697

Patent No. 6277974
 GENERAL INFORMATION:
 APPLICANT: COGENET NEUROSCIENCE, Inc.
 APPLICANT: Lo, Donald C.
 APPLICANT: Barney, Shawn
 APPLICANT: Thomas, Mary Beth
 APPLICANT: Portbury, Stuart D.
 APPLICANT: Puranam, Kasturi
 APPLICANT: Katz, Lawrence C.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
 TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
 TITLE OF INVENTION: CELL DEATH
 FILE REFERENCE: 10001-005-999
 CURRENT APPLICATION NUMBER: US/09/461,697
 CURRENT FILING DATE: 1999-12-14
 NUMBER OF SEQ ID NOS: 466
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 193
 LENGTH: 696
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-461-697-193

Query Match 0.9%; Score 51; DB 4; Length 696;
 Best Local Similarity 46.4%; Pred. No. 0.00017;
 Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

QY 2752 ggcaggagatgcaatgagagctcagagctgcacgcaaatcttatgataaacccgcg 2811
 Db 181 ggaagagaagaagaagaagcagctgacgacagaatgaataaagaagaagacacagaa 240
 QY 2812 gtagatgatgatgagagcgtgagaccttggagaagacacagaaagttcaaga 2871

```
Db 241 gaagatgaaagatcacaacgaagaagggagagctgtgaaagaagacaaagttaa 300
Oy 2872 tcaggagagagagctgtatccacagccagtgatcagcctccactctctgtgagagaca 2931
Db 301 aagagggaagagatggaagaagagataaaatgtgaataatgagaagagagatgtcaaaa 360
Oy 2932 atgaatgtaaaccaagagaaggtgtatgaatgtatgaataaaagataaaactgtgaaa 2991
Db 361 gagaagaagagatggaaaaaaggtg--aagacggaagaagaaatgtgagaagctgtgaaa 418
Oy 2992 gaaagaagaagatagatgaaggaaggaataaatgaaagccaaaggaagagagctgtc 3051
Db 419 agaaaggaagagatgtaaaagaaggaaggaagaaacaggaagctgtgaaaagaga 478
Oy 3052 aaggctgtggagacatgttcaggtttgtgcaaacatcgaaaagaatgacaaagtgtgaaa 3111
Db 479 atgaagatggaagaagagagaggaataaaagagggagaagatgtaaaagctcaagaag 538
Oy 3112 acgggtataataaaataacaggaatcctttacatcagaagaagagagatacgaatgaag 3171
Db 539 atgaaaagaagagagagatggaagaagaatgaaagtgtgaaatgtgagaagaagctgga 598
Oy 3172 caggagcagagaga 3184
Db 599 aagagaagaaga 611

RESULT 8
US-09-461-697-191
; Sequence 191, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 191
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-191

Query Match 0.9%; Score 51; DB 4; Length 699;
Best Local Similarity 46.4%; Pred. NO. 0.00018;
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;
```

```
Oy 2992 gaaagaagaagatagagataaggaagagataaaatgaaagccaaaggaagagatgtc 3051
Db 422 agaaaggaagagatgtaaaagaggaaggaagacagaaaagaacagaggtgtggaagaaga 481
Oy 3052 aaggctgtggagacatgttcaggtttgtgcaaacatcgaaaagaatgacaaagtgtgaaa 3111
Db 482 atgaagatggaaaaaagagaaggaaggaataaaagaaggggaagatgtgaaaagctcaagaag 541
Oy 3112 acgggtataataaaataacaggaatcctttacatcagaagaaggaaggaatgacgaatgaag 3171
Db 542 atgaaaagaagagagagagatggaagaagaatgaaagtgtgaaatgtgagaagaagcttggaa 601
Oy 3172 caggagcagagaga 3184
Db 602 aagagaagaaga 614

RESULT 9
US-09-461-697-189
; Sequence 189, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 189
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-189

Query Match 0.9%; Score 51; DB 4; Length 717;
Best Local Similarity 46.4%; Pred. NO. 0.00018;
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;
```

QY 3112 acgggttaataataataacaggaatccttaacatcagaagagagagagagagag 3171
Db 560 atgataaag 3171
QY 3172 caggagcaggagagag 3184
Db 620 aagagagagagagagag 632

RESULT 10

US-09-461-697-187
Sequence 187, Application US/09461697
Patent No. 6277974
GENERAL INFORMATION:
APPLICANT: COGENET NEUROSCIENCE, Inc.
APPLICANT: LO, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth
APPLICANT: Portbury, Stuart D.
APPLICANT: Putnam, Kasturi
APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461,697
NUMBER OF FILING DATE: 1999-12-14
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 187
LENGTH: 774
TYPE: DNA
ORGANISM: Homo sapiens
US-09-461-697-187

Query Match 0.9%; Score 51; DB 4; Length 774;
Best Local Similarity 46.4%; Pred. No. 0.00019;
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

QY 2752 ggcaggag 2811
Db 259 ggcaggag 2811
QY 2812 gtcag 318
Db 319 ggcaggag 2871
QY 2872 tcaggag 378
Db 379 aagggag 2931
QY 2932 atgag 438
Db 439 ggcaggag 2991
QY 2992 ggcaggag 496
Db 497 agaaag 3051
QY 3052 aagggag 556
Db 557 atgag 3111
QY 3112 acgggttaataataataacaggaatccttaacatcagaagagagagagagag 616
Db 617 atgataaag 3171
QY 3172 caggagcaggagagag 3184
Db 677 aagagagagagagagag 689

RESULT 11
US-09-461-697-185
Sequence 185, Application US/09461697
Patent No. 6277974
GENERAL INFORMATION:
APPLICANT: COGENET NEUROSCIENCE, Inc.
APPLICANT: LO, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth
APPLICANT: Portbury, Stuart D.
APPLICANT: Putnam, Kasturi
APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461,697
NUMBER OF FILING DATE: 1999-12-14
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 185
LENGTH: 819
TYPE: DNA
ORGANISM: Homo sapiens
US-09-461-697-185

Query Match 0.9%; Score 51; DB 4; Length 819;
Best Local Similarity 46.4%; Pred. No. 0.00019;
Matches 201; Conservative 0; Mismatches 230; Indels 2; Gaps 1;

QY 2752 ggcaggag 2811
Db 304 ggcaggag 363
QY 2812 gtcag 2871
Db 364 ggcaggag 423
QY 2872 tcaggag 483
Db 424 aagggag 2931
QY 2932 atgag 2991
Db 484 ggcaggag 541
QY 2992 ggcaggag 3051
Db 542 agaaag 601
QY 3052 aagggag 601
Db 602 atgag 3111
QY 3112 acgggttaataataataacaggaatccttaacatcagaagagagagagagag 661
Db 662 atgataaag 3171
QY 3172 caggagcaggagagag 3184
Db 722 aagagagagagagagag 734

RESULT 12
US-09-461-697-184
Sequence 184, Application US/09461697
Patent No. 6277974
GENERAL INFORMATION:
APPLICANT: COGENET NEUROSCIENCE, Inc.
APPLICANT: LO, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth


```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(7521)
; PUBLICATION INFORMATION:
; JOURNAL: FEBS Lett.
; VOLUME: 337
; ISSUE: 2
; PAGES: 200-206
; DATE: 1994-01-10
; DATABASE ACCESSION NUMBER: D21209/Genbank
; DATABASE ENTRY DATE: 1999-02-05
US-09-290-640-45

```

```

Query Match
Best Local Similarity 0.9%; Score 50.6; DB 4; Length 8119;
Matches 119; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

```

```

QY 1369 ggcagaaggcttaataatccagcttaagaaggtaacagaaggttgggaltccagctact 1428
DB 4153 ggaatatacttggagcttgacagctgcaaaaatgatacagcttgggataagtgtaacg 4212
QY 1429 tccagaagatgtaacaatagctgctcagctccaatctatgtgaaaaaatctctcccccg 1488
DB 4213 ggaagtgtaatacagaagtcacagacatggtgcatltagtaaaagctgtatcccccag 4272
QY 1489 ggggcggccaltcagaatggcgacttaagcgagagacagactatagaagtaaatgaa 1548
DB 4273 ggaagcgacagagctgtgtagaattcacaaggtgacgcgtccctagctgtcaatgaa 4332
QY 1549 gtagattagtggaacatcccaagaaggaagttgttcgctgttgagaagcac 1601
DB 4333 gtagcttagaaggacccacataagcaagctgtggaacacgtgagaataac 4385

```

RESULT 15

```

US-09-233-086-2
; Sequence 2, Application US/09233086
; Patent No. 6337192
; GENERAL INFORMATION:
; APPLICANT: Bartel, Paul L.
; APPLICANT: Taviqian, Sean V.
; TITLE OF INVENTION: MMS1 - An MMAC1 Interacting Protein
; FILE REFERENCE: MMS1 Gene
; CURRENT APPLICATION NUMBER: US/09/233,086
; EARLIER FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: US 60/071,861
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5836
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (115)..(5757)
US-09-233-086-2

```

```

Query Match
Best Local Similarity 0.9%; Score 49; DB 4; Length 5836;
Matches 124; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

```

```

QY 1381 aatccagcttaagaaggttaagaaggttggatcagcatcaccagagatgta 1440
DB 5509 attactacttgagaagaagctcctgaagcttgagggttagtatgtgagggttaagaa 5568
QY 1441 acataggttgctcagctcaatctatgtaaaacattctcccccggggcgccat 1500

```

```

DB 5569 agtcccatggagactgccaattatgtaacagctgattgcaaggagacagctgca 5628
QY 1501 caggatggccgaacttaaggcagaagacagactatagaagtaaatggaattagtg 1560
DB 5629 gatcagggccgattaaacagagggatcagattttagctgttaalggcgaacccctgaa 5688
QY 1561 ggcnaatcccaaggagagtgcttgcgttgagaagcacaagatggaaggaactgtg 1620
DB 5689 ggtgtactcatgagcaagcagtcgcattctaaacaccagagagagactgtaacctta 5748
QY 1621 agcctctg 1629
DB 5749 actgtgctg 5757

```

Search completed: July 24, 2002, 11:13:56
Job time: 15968 sec

QY 1863 caatgcatatataatgagagcagcatcctaagaatggaagcttcgggtgtaatgcatca 1922
DB 1692 caahtccattataatgagagcagcatcctaagaatggaagcttcgggtgtaatgcatca 1751
QY 1923 actatagaacataatgagaaatccctgttgggcaagaacaaacgaagtgcataatgaaac 1982
DB 1752 actatagaacataatgagaaatccctgttgggcaagaacaaacgaagtgcataatgaaac 1811
QY 1983 cctaagaagcgtatgtctactgtaaggaacataaagaaatgataccgcttatgtctgc 2042
DB 1812 cctaagaagcgtatgtctactgtaaggaacataaagaaatgataccgcttatgtctgc 1871
QY 2043 aagagaaataaagcaagtcaatgagctgaagctacccctggagagcccccctggaacatgagct 2102
DB 1872 aagagaaataaagcaagtcaatgagctgaagctacccctggagagcccccctggaacatgagct 1931
QY 2103 gcccaattgaacaaagcgtgtggtatgataagaagaaattcccaattccctcctaagaggg 2162
DB 1932 gcccaattgaacaaagcgtgtggtatgataagaagaaattcccaattccctcctaagaggg 1991
QY 2163 gattgaagggcttgatgaatacgcgccagagaataatgctgcctcctaagtaagtaatggtgca 2222
DB 1992 gattgaagggcttgatgaatacgcgccagagaataatgctgcctcctaagtaagtaatggtgca 2046
QY 2223 gtcaggttaataacagcgtgtcccttaacagtgaataatgcccagaatgtaacactgtcatat 2282
DB 2047 ---gttaataacagcgtgtcccttaacagtgaataatgcccagaatgtaacactgtcatat 2102
QY 2283 agaaataacagcgtgtcagcgtgtccctcctaactgtcagcagctgtcctcctaagctc 2342
DB 2103 agaaataacagcgtgtcagcgtgtccctcctaactgtcagcagctgtcctcctaagctc 2162
QY 2343 ccatgataatgttgggttcttgtaacgcagagatgctgtaacttgggcaagagcttgcataacga 2402
DB 2163 ccatgataatgttgggttcttgtaacgcagagatgctgtaacttgggcaagagcttgcataacga 2222
QY 2403 tgatcagcagcagcgtccttctgtaatcagaatgctgtaatccagcttctgtcttcaacagaa 2462
DB 2223 tgatcagcagcagcgtccttctgtaatcagaatgctgtaatccagcttctgtcttcaacagaa 2282
QY 2463 aggatttgaagctcagatgatagtcagaataacgcacaagaacatttcaagatgcagatca 2522
DB 2283 aggatttgaagctcagatgatagtcagaataacgcacaagaacatttcaagatgcagatca 2298
QY 2523 attgatttctgtaaaaaacagaaaaatacaaaaagcatgatttaagtataagctgcagagac 2582
DB 2299 -----atagctcagagagac 2312
QY 2583 taactcaatacagatgtaacagagaagcaggttctccacagagagatggtgtccttc 2642
DB 2313 taactcaatacagatgtaacagagaagcaggttctccacagagagatggtgtccttc 2372
QY 2643 cctgggtctgagaagatcaagctcttgagaagcttgcagaaacagagatgcagaaagtgcac 2702
DB 2373 cctgggtctgagaagatcaagctcttgagaagcttgcagaaacagagatgcagaaagtgcac 2432
QY 2703 ttggaatggggaatatctcttcacatcgtccacagcgcgaggaataacagagcagggagatg 2762
DB 2443 ttggaatggggaatatctcttcacatcgtccacagcgcgaggaataacagagcagggagatg 2492
QY 2763 caatgaagcgttcagagcttcagcatcgcacaatactatgataaaacccggtgataatgata 2822
DB 2493 caatgaagcgttcagagcttcagcatcgcacaatactatgataaaacccggtgataatgata 2552
QY 2823 tgatgaagcgttgagacgttcgagaagagacacagaagaagtccaagaatcagagagagata 2882
DB 2553 tgatgaagcgttgagacgttcgagaagagacacagaagaagtccaagaatcagagagagata 2612
QY 2883 gtctgtatccacagcagtgatcagcttccacactctcggagagacaataatgataatgataa 2942
DB 2613 gtctgtatccacagcagtgatcagcttccacactctcggagagacaataatgataatgataa 2672
QY 2943 ccaagaagaaggtgataagactgatagaataaagaataaactgtaaaaaagaagaaga 3002

DB 2673 ccaagaagaaggtgataagactgatagaataaagaataaactgtaaaaaagaagaaga 2732
QY 3003 agatagaataagagaaagatataaataagaaacgaagaaggaatgctgaagggcttggg 3062
DB 2733 agatagaataagagaaagatataaataagaaacgaagaaggaatgctgaagggcttggg 2792
QY 3063 agacatgttcagagtttggcaaaacatcgaaaagaatgacaagatttgaagaaacgggtlaaat 3122
DB 2793 agacatgttcagagtttggcaaaacatcgaaaagaatgacaagatttgaagaaacgggtlaaat 2852
QY 3123 aaaaatacagaagaaatcccttacaatcagaagaaggaaggaatgaaatgaaacagagaa 3182
DB 2853 aaaaatacagaagaaatcccttacaatcagaagaaggaaggaatgaaatgaaacagagaa 2912
QY 3183 gaggatctcaagccaaaacacgcagaaatttgaagaaacagacgtctgagaaacgtgataatgc 3242
DB 2913 gaggatctcaagccaaaacacgcagaaatttgaagaaacagacgtctgagaaacgtgataatgc 2972
QY 3243 tgaataatcaagaatttcalcgagacatttggctgtgataatgataatgataatgataatgata 3302
DB 2973 tgaataatcaagaatttcalcgagacatttggctgtgataatgataatgataatgataatgata 3032
QY 3303 tctctcttaatgaaggttccatgctcacaagcctgagacctcagagccccaagaaaggaca 3362
DB 3033 tctctcttaatgaaggttccatgctcacaagcctgagacctcagagccccaagaaaggaca 3092
QY 3363 tatgataatgagcttctgataatgagcccaagtcagaagccggagaaattccaaacccctcact 3422
DB 3093 tatgataatgagcttctgataatgagcccaagtcagaagccggagaaattccaaacccctcact 3152
QY 3423 agacagataacagatacaactcctcagaaatcatcagataatcagatcagatcagatcagatc 3482
DB 3153 agacagataacagatacaactcctcagaaatcatcagataatcagatcagatcagatcagatc 3212
QY 3483 tcaagcaagcaagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 3542
DB 3213 tcaagcaagcaagaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 3272
QY 3543 accctggccgaagcgcagc 3602
DB 3273 accctggccgaagcgcagc 3332
QY 3603 gcaagatgcagc 3662
DB 3333 gcaagatgcagc 3392
QY 3663 cagctctcgc 3722
DB 3393 cagctctcgc 3452
QY 3723 gcaagactatccctcggggaagagcttccagaggtgcgaagaagaaacccaggttaactccag 3782
DB 3453 gcaagactatccctcggggaagagcttccagaggtgcgaagaagaaacccaggttaactccag 3512
QY 3783 ctacaaggtctccaggaacgcgtacccctggagagacatggtcttcaacgcgaaggtcatgct 3842
DB 3513 ctacaaggtctccaggaacgcgtacccctggagagacatggtcttcaacgcgaaggtcatgct 3572
QY 3843 ggaactcagagagcttccctcgcgcgaagaagaagagcgcgaagagcgcgaagatgaaagaaga 3902
DB 3573 ggaactcagagagcttccctcgcgcgaagaagaagagcgcgaagagcgcgaagatgaaagaaga 3632
QY 3903 gctccttcgagggggccagcaacatgataatgaaagatccagagcccccgtta 3962
DB 3633 gctccttcgagggggccagcaacatgataatgaaagatccagagcccccgtta 3692
QY 3963 ggcctcccaaggggccccttcgcgaagatggtcccccctcccttcaaggttgcag 4022
DB 3693 ggcctcccaaggggccccttcgcgaagatggtcccccctcccttcaaggttgcag 3752
QY 4023 gctgaacagacttcagacctcctgagaagaagggcccttctatctctgagcagcaata 4082

```

Db 3753 gctgaacagacttcagactcctcgtgagaagggagcccttcattctcctgagcaacgaata 3812
Oy 4083 acggatgcttcacatgctgcgcgaataaagaacatttccctatggaagctgttattttggag 4142
Db 3813 acggatgcttcacatgctgcgcgaataaagaacatttccctatggaagctgttattttggag 3872
Oy 4143 ttttttaaacctcgatgctacatgagatattctgtttgtttgatalcagtgctttaa 4202
Db 3873 ttttttaaacctcgatgctacatgagatattctgtttgtttgatalcagtgctttaa 3932
Oy 4203 gcggtgtaagcaaaagaaaggaagccttaagtcttcttgcacactgtgtcgaagtgctg 4262
Db 3933 gcggtgtaagcaaaagaaaggaagccttaagtcttcttgcacactgtgtcgaagtgctg 3992
Oy 4263 ttcatgagaagatctccacactctgacaacatcattgtttgaggtgttcatalgtctg 4322
Db 3993 ttcatgagaagatctccacactctgacaacatcattgtttgaggtgttcatalgtctg 4052
Oy 4323 cgcctctcacagatcccaagaaatcgcgcctacacagatgtgtgcgcgctgtgtgt 4382
Db 4053 cgcctctcacagatcccaagaaatcgcgcctacacagatgtgtgtgcgcgctgtgtgt 4112
Oy 4383 aacatccctgcacacactgtgcagtgacaatctcaactgagaatgagaagtgagctgcgc 4442
Db 4113 aacatccctgcacacactgtgcagtgacaatctcaactgagaatgagaagtgagctgcgc 4172
Oy 4443 tgttctcccttaagctctcttctatgataagatgacatttaagctctgttggaggaaa 4502
Db 4173 tgttctcccttaagctctcttctatgataagatgacatttaagctctgttggaggaaa 4232
Oy 4503 agtgggacatacacaaaatttgggcttctgtgtcttcaagccttgcagcaacacattgtc 4562
Db 4233 agtgggacatacacaaaatttgggcttctgtgtcttcaagccttgcagcaacacattgtc 4292
Oy 4563 gtaacttgcactactgtgtgttcttcttcttctccttcaagccttgcagggagagatc 4622
Db 4293 gtaacttgcactactgtgtgttcttcttcttctccttcaagccttgcagggagagatc 4352
Oy 4623 ggtacatccattagaagaagccttaagccttaagccttaagccttgcacatggagggtcc 4682
Db 4353 ggtacatccattagaagaagccttaagccttaagccttaagccttgcacatggagggtcc 4412
Oy 4683 tttgttaactgccttaagatgactcaacatcaacataaagaatgagtgatctgtctgtt 4742
Db 4413 tttgttaactgccttaagatgactcaacatcaacataaagaatgagtgatctgtctgtt 4472
Oy 4743 taacatagctgttcggttccgttccctgcgaacgtgggttgatgacacacagctctgtg 4802
Db 4473 taacatagctgttcggttccgttccctgcgaacgtgggttgatgacacacagctctgtg 4532
Oy 4803 gggagacgtgggagacagagtggtctgcgcgggggagagctgggcttcgacattggccctg 4862
Db 4533 gggagacgtgggagacagagtggtctgcgcgggggagagctgggcttcgacattggccctg 4592
Oy 4863 tgtcatcatgaggggagagctagaagaagaattctcctcgaagaagctatggccagt 4922
Db 4593 tgtcatcatgaggggagagctagaagaagaattctcctcgaagaagctatggccagt 4652
Oy 4923 acatccatgaatattatttaattagtttctgtctgcagcctgtgcaaggagacaga 4982
Db 4653 acatccatgaatattatttaattagtttctgtctgcagcctgtgcaaggagacaga 4712
Oy 4983 atgggacagatataaacacacagatcatttgaatcgtcttca-cggttttcaagta 5041
Db 4713 atgggacagatataaacacacagatcatttgaatcgtcttca-cggttttcaagta 4772
Oy 5042 gaggcaatgatagtctgtcctaataatgacacacacatgcatgtctgactggcgatgcaag 5101
Db 4773 gaggcaatgatagtctgtcctaataatgacacacacatgcatgtctgactggcgatgcaag 4832
Oy 5102 ctccatagatgtctgcctgaaacataaagaacaaaggaaaagccgttaacatcacac 5161
Db 4833 ctccatagatgtctgcctgaaacataaagaacaaaggaaaagccgttaacatcacac 4892

```

```

Oy 5162 agagacaatttcgggtccacacagcgtgtgtgaggaagaagctacatctcgcgtcagat 5221
Db 4893 agagacaatttcgggtccacacagcgtgtgtgaggaagaagctacatctcgcgtcagat 4952
Oy 5222 agagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5281
Db 4953 agagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5012
Oy 5282 catctgtatgttaaacacatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5341
Db 5013 catctgtatgttaaacacatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5072
Oy 5342 aagctggaggtgaaagcttcagatcagatgatttactttaaaacagaaaagaacat 5401
Db 5073 aagctggaggtgaaagcttcagatcagatgatttactttaaaacagaaaagaacat 5132
Oy 5402 gtagatgatagtctatatttttttaattggacatgtattttgtgtgtgtgtgtgt 5461
Db 5133 gtagatgatagtctatatttttttaattggacatgtattttgtgtgtgtgtgtgt 5192
Oy 5462 agaaatgatagtctcacaacacagtaaccgtgtctctcctgacattctgtgcatgtctg 5521
Db 5193 agaaatgatagtctcacaacacagtaaccgtgtctctcctgacattctgtgcatgtctg 5252
Oy 5522 ttcttaatacgtgcggcggtgtctaaagtgtgttaacagtgtaagcgaatgacattg 5581
Db 5253 ttcttaatacagtggtgcggcggtgtctaaagtgtgttaacagtgtaagcgaatgacattg 5312
Oy 5582 gatgacagtggtctcttcttcaacagcctccctgagctgtgagaagaagcttctctgac 5641
Db 5313 gatgacagtggtctcttcttcaacagcctccctgagctgtgagaagaagcttctctgac 5372
Oy 5642 atatgaacaccttaataaagaacatatttctcctgtttaaaaaa 5688
Db 5373 atatgaacaccttaataaagaacatatttctcctgtttaaaaaa 5419

RESULT 2
AAH18587
ID AAH18587 standard; cDNA; 2718 BP.
XX
AC AAH18587;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18774.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection

```

PT and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs - PT

PS Claim 8; SEQ ID 18774; 2537pp + CD ROM; English..

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-3T primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

SQ Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other;

Query Match	Score	DB	Length
Best Local Similarity	2476.8;	22;	2718;
Matches 2603; Conservative	96.08;		
	Pred. No. 0;		
	Mismatches 7;		
	Indels 102;		
	Gaps 2		

OY	374	ttgaggtcaacaccttcagtccttcgagcaaatatgctcttcattcatgltcgagcagtagtg	433
Db	1	ttgaggtcaacaccttcagtccttcgagcaaatatgctcttcattcatgltcgagcagtagtg	60
OY	434	accgaagctcattatgggcctccacacctctctgacgtatgataattttctctcgaagagc	493
Db	61	accgaagctcattatgggcctccacacctctctgacgtatgataattttctctcgaagagc	120
OY	494	cttcaagagaaaaatcccaacacgctgtgtcaacaacagcttgcctctcaagcagaacactg	553
Db	121	cttcaagagaaaaatcccaacacgctgtgtcaacaacagcttgcctctcaagcagaacactg	180
OY	554	cttgaggagtcctcaaaaacctcgcgacagcaggaagaagttgaaaaactcagagagctccggcg	613
Db	181	cttgaggagtcctcaaaaacctcgcgacagcaggaagaagttgaaaaactcagagagctccggcg	240
OY	614	atactagtgtaacttggcttctaacaacatttcagaagagacaaatgtctgcgtctctcgtatgtcca	673
Db	241	atactagtgtaacttggcttctaacaacatttcagaagagacaaatgtctgcgtctctcgtatgtcca	300
OY	674	gtccaccaatgtgtggtgcgaatggtgtctgagaagaaacagacagatgagatgaggaacggaag	733
Db	301	gtccaccaatgtgtggtgcgaatggtgtctgagaagaaacagacagatgagatgaggaacggaag	360
OY	734	aggatataaagtcggtgttaaacctgtctgagacatgtctgacacgggttttgagacatatacca	793
Db	361	aggatataaagtcggtgttaaacctgtctgagacatgtctgacacgggttttgagacatatacca	420
OY	794	actttctctgtagatatagtgtaaagctcgtatagaatgccccaacgatgtagagagcctctg	853
Db	421	actttctctgtagatatagtgtaaagctcgtatagaatgccccaacgatgtagagagcctctg	480
OY	854	gaatccatgtatgaccttcttaagtgactcgaagcggcacaacacctggggcttatcttgataaac	913
Db	481	gaatccatgtatgaccttcttccacgtctctcgaagcggcagaacctggggcttatcttgataaac	540
OY	914	gatttgagaagaagtggttaaaactgtacaatgtaaaaactcttttcgttgataatgatattgacttg	973

|||||
Db 121 ctccaagaaataacccacagctggtcaacaacagctcgcttccctcaagcaactcgt 180
QY 554 ctggagagcttaaaaaacctgcgcacagaaagaaatgaaataactacaagaagcctccgcg 613
Db 181 ctggagagcttaaaaaacctgcgcacagaaagaaatgaaataactacaagaagcctccgcg 240
QY 614 ataactagtaactggtcttaaccaattctcagagagacaatgctcgtctcgtctagtgcca 673
Db 241 ataactagtaactggtcttaaccaattctcagagagacaatgctcgtctcgtctagtgcca 300
QY 674 gtaccccaatggtgggcagatggcggtgcgagaaacaagaacagatagatggagagaaag 733
Db 301 gtaccccaatggtgggcagatggcggtgcgagaaacaagaacagatagatggagagaaag 360
QY 724 aggaataaagctggtcttaaacctgtctgacatgctgcacagcgggtctggagcaataacca 793
Db 361 aggaataaagctggtcttaaacctgtctgacatgctgcacagcgggtctggagcaataacca 420
QY 794 actttctctggaatgaatggttaagaagctgcgaagtcaccaagatggagggcctcgtg 853
Db 421 actttctctggaatgaatggttaagaagctgcgaagtcaccaagatggagggcctcgtg 480
QY 854 gaatccatgtagtgcttcagtgctgcagagcgagaaacccctggggttatatagtaaac 913
Db 481 gaatccatgtagtgcttcagtgctgcagagcgagaaacccctggggttatatagtaaac 540
QY 914 gattggagaaaggttgtaaaagctgcacatgtaaaactctttcgtatagaatgattgcatg 973
Db 541 gattggagaaaggttgtaaaagctgcacatgtaaaactctttcgtatagaatgattgcatg 600
QY 974 tcaagataatgatagtagcagccttcgaaatagaagaattgtaacaagcacacaatagtttc 1033
Db 601 tcaagataatgatagtagcagccttcgaaatagaagaattgtaacaagcacacaatagtttc 660
QY 1034 gccaagccaatgctgtaacacacacatcttggttccatggttccctgcagcaataaagagc 1093
Db 661 gccaagccaatgctgtaacacacacatcttggttccatggttccctgcagcaataaagagc 720
QY 1094 agtatgaaacaactacccaagaatgagaagaacaattactactcaagccgttttaagccgt 1153
Db 721 agtatgaaacaactacccaagaatgagaagaacaattactactcaagccgttttaagccgt 780
QY 1154 acagccagatatactgacaaacagagatgtagaacaagtgcagggcttcaacagtgacagag 1213
Db 781 acagccagatatactgacaaacagagatgtagaacaagtgcagggcttcaacagtgacagag 840
QY 1214 caccgcgactggaacacacccgcttgagcagatagactctcaactcaagactcaactcaagc 1273
Db 841 caccgcgactggaacacacccgcttgagcagatagactctcaactcaagactcaactcaagc 900
QY 1274 caaacacccctcgggaaacaacacacatcgcgtccagcctcgcagcactcaagatglatatagta 1333
Db 901 caaacacccctcgggaaacaacacacatcgcgtccagcctcgcagcactcaagatglatatagta 960
QY 1334 cgaactgtaagcagtggtataaacaacaaataagccaagaggtcttaataccagctta 1393
Db 961 cgaactgtaagcagtggtataaacaacaaataagccaagaggtcttaataccagctta 1020
QY 1394 agaaaggtgacagaaggtttggatccagatcaactccagagatgtaaatagttggt 1453
Db 1021 agaaaggtgacagaaggtttggatccagatcaactccagagatgtaaatagttggt 1080
QY 1454 cagctcccaatctatgtgaaaaaacaattctcccccgggggcgcacatccagaatgagccagc 1513
Db 1081 cagctcccaatctatgtgaaaaaacaattctcccccgggggcgcacatccagaatgagccagc 1140
QY 1514 ttaaggtcagagacagactatacagagtaaatgtagattagttgggccaatcccaag 1573
Db 1141 ttaaggtcagagacagactatacagagtaaatgtagattagttgggccaatcccaag 1200
QY 1574 aggaaggtgttctcgtgtgtagaagcaacaagatgaaaggaactgtgaagccttcgtgtc 1633
|||||

Db 1201 aggaaggtgttctcgtgtgtagaagcaccaagaatgaaaggaactgtgaagccttcgtgtc 1260
QY 1634 ttcgcagagaaagcgccttccacccaaggaactgaaatgacagacgaagccagatgacaga 1693
Db 1261 ttcgcagagaaagcgccttccacccaaggaactgaaatgacagacgaagccagatgacaga 1320
QY 1694 ttccaaaagaaacgaagacagaatgagaatattgtcttcaacacctgataggcaccaag 1753
Db 1321 ttccaaaagaaacgaagacagaatgagaatattgtcttcaacacctgataggcaccaag 1380
QY 1754 aattctgaatattgaagcccaactaaatgatacagagatcgcagggccttggtgtcaggt 1813
Db 1381 aattctgaatattgaagcccaactaaatgatacagagatcgcagggccttggtgtcaggt 1440
QY 1814 tcaaaagtaaccgggtcaaaaagaaaccaacgcagatcttggaatcctttgcaagtcocat 1873
Db 1441 tcaaaagtaaccgggtcaaaaagaaaccaacgcagatcttggaatcctttgcaagtcocat 1500
QY 1874 ttaatggaagagcagacatcaaaagatggaaggtctcgggtgaaatgatacgaatgacag 1933
Db 1501 ttaatggaagagcagacatcaaaagatggaaggtctcgggtgaaatgatacgaatgacag 1560
QY 1934 taaatggaagatccctgttggtgcagacaaacaagaatgccaatggaacccctaagaaggt 1993
Db 1561 taaatggaagatccctgttggtgcagacaaacaagaatgccaatggaacccctaagaaggt 1620
QY 1994 ctatgtctataagagcaataaaagagaaatgatacgaacttttgttgcaagagaaata 2053
Db 1621 ctatgtctataagagcaataaaagagaaatgatacgaacttttgttgcaagagaaata 1680
QY 2054 gcaagtgcaatgagctgaaagtcacctgtagagcccccctggaactgagctgacatgaa 2113
Db 1681 gcaagtgcaatgagctgaaagtcacctgtagagcccccctggaactgagctgacatgaa 1740
QY 2114 cagcgtttgataatgagaagaacgaagaattcccatctcccttaacagttgaggtgagggc 2173
Db 1741 cagcgtttgataatgagaagaacgaagaattcccatctcccttaacagttgaggtgagggc 1800
QY 2174 tlgatgaaatcgccccaagcaaaatgctgcgcctcagatgagaataatggtgagcaagtaat 2233
Db 1801 tlgatgaaatcgccccaagcaaaatgctgcgcctcagatgagaataatggtgagcaagtaat 1851
QY 2234 accagctgtccctcagatgaaatgagcccaagatgacaatgcatatagaagaatgaca 2293
Db 1852 accagctgtccctcagatgaaatgagcccaagatgacaatgcatatagaagaatgaca 1911
QY 2294 ggttgccaatggttctctccaacatctctcgacaagcccttcacagatcccatgatatg 2353
Db 1912 ggttgccaatggttctctccaacatctctcgacaagcccttcacagatcccatgatatg 1971
QY 2354 tggggtttgtgacggcagaatgctggtacttgggccaaggtctgcaatcagtgatccagcg 2413
Db 1972 tggggtttgtgacggcagaatgctggtacttgggccaaggtctgcaatcagtgatccagcg 2031
QY 2414 actgctctttgagttccaagatggttgatccaagttccttgcttcaacgaagaagaattggac 2473
Db 2032 actgctctttgagttccaagatggttgatccaagttccttgcttcaacgaagaagaattggac 2091
QY 2474 gtcaagatgtagcaagaanaacgaanaagcaatttcagatgccaatgaaattgattcg 2533
Db 2092 gtcaaga----- 2097
QY 2534 ttaaaaacgaanaatacaaaaacgatgattagtaatacgtacagagacataaata 2593
Db 2098 -----ctgacggactaaatacaata 2118
QY 2594 cagtggaatgacaagaagaaggttctccacgacagagaatgttggttcccttcgtggtctga 2653
Db 2119 cagtggaatgacaagaagaaggttctccacgacagagaatgttggttcccttcgtggtctga 2178
QY 2654 agaaagtcaggtcgtgtggagatctgcagacccgaattccgaaggtgacattgaaatgggg 2713
Db 2179 agaaagtcaggtcgtgtggagatctgcagacccgaattccgaaggtgacattgaaatgggg 2238
|||||

OY	2714	attctcttcacgtgtccagcgcccgagataatcagagcgaggtatgcatgagact	2773
Db	2239	atatctctttccatcgcgcacgycgcgcgcatatccagagccagggatgcgaatgagact	2298
OY	2774	tcagagctgcgcatacgcagacaactcttatgataaacccgcggtlaatatgatagtatgtagagca	2833
Db	2239	tcagagctgcgcatacgcagacaactcttatgataaacccgcggtlaatatgatagtatgtagagca	2358
OY	2834	tggagaccttggagaagaaagacacagaaagaaagtctcaagatacagggagagagctgtatcca	2893
Db	2359	tggagaccttggagaagaaagacacagaaagaaagtctcaagatacagggagagagctgtatcca	2418
OY	2894	cagcgacgtgatcagccttcccaactctctcggagagacaaatgaaatgaaacccaagagaaag	2953
Db	2419	cagcgacgtgatcagccttcccaactctctcggagagacaaatgaaatgaaacccaagagaaag	2478
OY	2954	gtgataagactgtatagaaaaaaggaataaactcgtgaaaagaaaagaaagaaagatatagagata	3013
Db	2479	gtgataagactgtatagaaaaaaggaataaactcgtgaaaagaaaagaaagaaagatatagagata	2538
OY	3014	agaggaagagataaaatgaaagccaagaaagagaaatgcttgaagagcttgggagacatggttcca	3073
Db	2539	agaggaagagataaaatgaaagccaagaaagagaaatgcttgaagagcttgggagacatggttcca	2598
OY	3074	ggtttggaacac 3085	
Db	2599	gacctgccaacac 2610	

```

Result# 4
AAD01182
ID AAD01182 standard; cDNA; 2554 BP.
XX
XX
AC AAD01182:
XX
XX 04-OCT-2000 (first entry)
DT
DE PDZ domain-comprising PHIP (ephrin Interacting Protein) complex cDNA.
XX
XX PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
XX cellular process; nerve cell interaction; regeneration of nerve cell;
XX axonogenesis; antiproliferative; proliferative disorder; treatment; GRP;
XX differentiative disorder; PHIP; ephrin interacting protein; syntenin; ss
XX
OS Mus sp.
XX
XX key Location/Qualifiers
FH 306..2540
CDS /*tag= a
/*product= "PHIP (ephrin Interacting Protein) complex"
/note= "Homologous to C. elegans PAR-3 protein"
FT
FT
FT
XX
XX PN
XX
XX 02-JUN-2000.
XX
XX 19-NOV-1999; 99WC-CA01101.
XX PF
XX 20-NOV-1998; 98US-0109158.
XX PR
XX (MOUN ) MOUNT SINAI HOSPITAL.
PA
XX
XX Lin D, Pawson A;
PI
XX WPI; 2000-400038/34.
DR P-PSDB; AA171410.
XX
XX Isolated complex for treating proliferative or differentiative
PT disorders comprises B class ephrin and PDZ domain containing protein -
XX
XX disclosure; page 57-59; 59pp; English.
XX

```

Score 1695.8; DB 21; Length 2554;

Query Match	25.8%	Pred. No. 0;	Indels 0;	Gaps 0
Query Similarity	84.8%	Mismatches 342;		
Best Local Similarity	0;			
Matches 1901; Conservative				

QY	1	atgaagatgacccctgtgcttccgaggaagaccgcgggtggtctgtgcgcgttcgcggagaaatgtagcgc	300
Db	306	atgaagatgacccctgtgcttccgaggaagaccgcgggtggtctgtgcgcgttcgcggagaaatgtagcgc	120
QY	61	atgaagaatttcagaccgccatccacagcagcggtgacccgcgtaccacccgaaagccatccgcaag	425
Db	366	atgaagaatttcagaccgccatccacagcagcggtgacccgcgtaccacccgaaagcgcggtgagcaag	425
QY	121	gatcccaactctgtgatacaagctgacatcgctcttggaacaatggaagatggagaggaatctagac	180
Db	426	gatcccaactctgtgatacaagctgacatcgctcttggaacaatggaagatggagaggaatctagac	485
QY	181	ctgtatgatcatctcttctgtgtatgtagcagacgaataaagaacgagactgtgtacagtgttat	240
Db	486	ctgtatgatcatctcttctgtgtatgtagcagacgaataaagaacgagactgtgtatgaagatattgat	545
QY	241	gagcagatccacacatccacgagagtgatgagccacgagctccagctccacgctccacggtacccagac	300
Db	546	gagcagatccacacatccacgagagtgatgagccacgagctccagctccacggtacccagacccagat	605
QY	301	ccagagatatttgtagatgtagagctctggacccaacaatgtctgaagcttccagcttccaccca	360
Db	606	ccagagatatttgtagatgtagagctctggacccaacaatgtctgaagcttccagcttccaccca	665
QY	361	gcaacaagatgaatatgtagagctccacactcagctcagcttcgcagcaatatatgccccttcaagtt	420
Db	666	gcaacaagatgaatatgtagagctccacactcagctcagcttcgcagcaatatatgccccttcaagtt	480
QY	421	cagcagatgtagtgcacacgtccacactcagctcagcttcgcagcaatatatgccccttcaagtt	725
Db	726	cagcagatgtagtgcacacgtccacactcagctcagcttcgcagcaatatatgccccttcaagtt	785
QY	481	tctctgtagagacgtcttcacagaaataatcccaacagctcgtgccaacaacagctgtgctctc	540
Db	786	tctctgtagagacgtcttcacagaaataatcccaacagctcgtgccaacaacagctgtgctctc	845
QY	541	aagcagaaacatctctgtgaggtcctctaaaacctcgtgcacaggaagaagatgtaaaactacaga	600
Db	846	aagcagaaacatctctgtgaggtcctctaaaacctcgtgcacaggaagaagatgtaaaactacaga	905
QY	601	agcctccgcgggagatctgtaactctgtgtcaacaaatttcagagagacaatgctgcgtcg	660
Db	906	agcctccgcgggagatctgtaactctgtgtcaacaaatttcagagagacaatgctgcgtcgctcc	965
QY	661	tctctgagtgcaatccccaactggtggtgcgaatgtctgagaaacaagaaacagatagag	720
Db	966	tctctgagtgcaatccccaactggtggtgcgaatgtctgagaaacaagaaacagatagag	1025

CC cervical cancer or has a pre-malignant condition; to monitor the
 CC progression of cervical cancer or a premalignant condition in a patient;
 CC and to select and/or assess the efficacy of a compound or therapy for
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be
 CC useful for gene therapy.
 XX

SQ Sequence 1800 BP; 578 A; 359 C; 466 G; 394 T; 3 other;

Query Match 28.8%; Score 1638.6; DB 22; Length 1800;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 1660; Conservative 0; Mismatches 4; Indels 9; Gaps 1;

QY 1413 ggatcagcatcactccagagatgttaacaataagtgctcagctccatctatgttaa 1472
 Db 3 ggattcagatcactccagagatgttaacaataagtgctcagctccatctatgttaa 1472
 QY 1473 aaacattctcccccggggggggccatcagatgagcagacttaaggcagagagagact 1532
 Db 63 aaacattctcccccggggggggccatcagatgagcagacttaaggcagagagagact 1532
 QY 1533 tataaggttaaatgaggtatgttggcgaatcccaagaagagtgctgtcgtgt 1592
 Db 123 tataaggttaaatgaggtatgttggcgaatcccaagaagagtgctgtcgtgt 1592
 QY 1593 gagaagcaccaagaatggaagaaactgtgagccttctgtgttcttgcagagaaagcgtt 162
 Db 183 gagaagcaccaagaatggaagaaactgtgagccttctgtgttcttgcagagaaagcgtt 1652
 QY 1653 ccaccacaaggaaactgcatctgacagcccaagcagatcccaaaaagaacgaagc 1712
 Db 243 ccaccacaaggaaactgcatctgacagcccaagcagatcccaaaaagaacgaagc 1712
 QY 1713 aagaagatagatatgttcttaccctgtatggcacaaggaattcttaccatttgaagt 1772
 Db 303 aagaagatagatatgttcttaccctgtatggcacaaggaattcttaccatttgaagt 1772
 QY 1773 cccacttaatgatctcagatctgcagacgttggctgctgcaagtgtcaaaaggttaacgttcaaa 1832
 Db 363 cccacttaatgatctcagatctgcagacgttggctgctgcaagtgtcaaaaggttaacgttcaaa 1832
 QY 1833 agaaacacagcagatcttggaactcttgcagatccatattatattgaagagacagc 1892
 Db 423 agaaacacagcagatcttggaactcttgcagatccatattatattgaagagacagc 1892
 QY 1893 taaagatggaaggtctcggtgaatgatcaatgtatagcagtaaatggaagatccctgtt 1952
 Db 483 taaagatggaaggtctcggtgaatgtatagcagtaaatggaagatccctgtt 1952
 QY 1953 gggcagaagaacacagatctggaactcttgcagatccatattatattgaagagacagc 2012
 Db 543 gggcagaagaacacagatctggaactcttgcagatccatattatattgaagagacagc 2012
 QY 2013 taaacgaggaatgatcagcttattgttgcagaaggaataagcaagtgcacatgtcaa 2072
 Db 603 taaacgaggaatgatcagcttattgttgcagaaggaataagcaagtgcacatgtcaa 2072
 QY 2073 gtccacttggaagccccccttgagcctgagctgtccatgtgaacagcgttgcagataga 2132
 Db 663 gtccacttggaagccccccttgagcctgagctgtccatgtgaacagcgttgcagataga 2132
 QY 2133 acgagaattcccatcctctacagtggtgaggttggaggtcttgatgaatgcgccagag 2192
 Db 723 acgagaattcccatcctctacagtggtgaggttggaggtcttgatgaatgcgccagag 2192
 QY 2193 aaatgctgcctcagtagagataagtgagtcaggttaataaccagcgttccccaagt 2252
 Db 783 aaatgctgcctcagtagagataagtgagtcaggttaataaccagcgttccccaagt 2252
 QY 2253 gaatagcccccaagatgacatgtcatatagaagaatacaggttgcagtgctctcc 2312
 Db 834 gaatagcccccaagatgacatgtcatatagaagaatacaggttgcagtgctctcc 2312

QY 2313 acatctctgacacagtcctcctccagcctccacatgatgctggtgttgcagcgaga 2372
 Db 894 acatctctgacacagtcctcctccagcctccacatgatgctggtgttgcagcgaga 2372
 QY 2373 tgcctgacttggtgccaaggtctgcaatcagtgatctcagcagcgtcttggatccaga 2432
 Db 954 tgcctgacttggtgccaaggtctgcaatcagtgatctcagcagcgtcttggatccaga 2432
 QY 2433 tgtttaccagttctgtcttccacaggaaggttggagctgatagttcagaaaaa 2492
 Db 1014 tgtttaccagttctgtcttccacaggaaggttggagctgatagttcagaaaaa 2492
 QY 2493 acgccaagaagcattttcagatgcacgttcaatttggatctgtttaaaccgaaaaa 1073
 Db 1074 acgccaagaagcattttcagatgcacgttcaatttggatctgtttaaaccgaaaaa 1073
 QY 2553 aagcatgatttagtatagtcagcagactaaactcaatacagtgatgaccagaagc 1133
 Db 1134 aagcatgatttagtatagtcagcagactaaactcaatacagtgatgaccagaagc 1133
 QY 2613 agttctcccgagcagaaatgtggtctctccctggtctcgaagaagtaagctgttga 1193
 Db 1194 agttctcccgagcagaaatgtggtctctccctggtctcgaagaagtaagctgttga 2672
 QY 2673 gattctgcagacagcagttgcagaggttgacttgaattgagatcttccatctgtcc 1253
 Db 1254 gattctgcagacagcagttgcagaggttgacttgaattgagatcttccatctgtcc 2732
 QY 2733 acgcccggagataatagagcagagggatgcaatgaggtctcagagctcagtcagca 2792
 Db 1314 acgcccggagataatagagcagagggatgcaatgaggtctcagagctcagtcagca 1373
 QY 2793 attctatgataaccgcggtgatgatgatgataaggaagcattggaagaaga 2852
 Db 1374 attctatgataaccgcggtgatgatgatgataaggaagcattggaagaaga 1433
 QY 2853 cacagaagaagttcaagaataaggaaggaagctgtatccacagcagtgatcaagctt 2912
 Db 1434 cacagaagaagttcaagaataaggaaggaagctgtatccacagcagtgatcaagctt 1493
 QY 2913 cactctctgagagacaaatgaaatggaagcagaaggaagtgataaagactgtatgaa 2972
 Db 1494 cactctctgagagacaaatgaaatggaagcagaaggaagtgataaagactgtatgaa 1553
 QY 2973 aaagataaaacttgaaagaagaagaagatagatagaataaggaaggaatgaa 3032
 Db 1554 aaagataaaacttgaaagaagaagaagatagatagaataaggaaggaatgaa 1613
 QY 3033 agccaagaaggaatgctgaaggtcttggaacatgttcaaggttggcaaac 3085
 Db 1614 agccaagaaggaatgctgaaggtcttggaacatgttcaaggttggcaaac 1666

RESULT 6
 AAF91875 standard; cDNA; 867 BP.
 XX AAF91875;
 AC
 XX
 XX
 DT 22-MAY-2001 (first entry)
 XX
 DE Human secreted protein-encoding gene 18 cDNA clone HPEFK57, SEQ ID NO:28.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW cardiovascular disorder; sepsis; diabetes; kidney disorder;
 KW gastrointestinal disorder; angiogenic disorder; kidney disorder;
 KW endocrine disorder; infection; pregnancy-related disorder;
 KW cell culture; chemotaxis; food additive;

KW binding partner identification: ss.
XX Homo sapiens.
OS
XX WO200118022-A1.
PN
XX 15-MAR-2001.
PD
XX 31-AUG-2000; 2000WO-US24008.
PF
XX 03-SEP-1999; 99US-0152315.
PR 03-SEP-1999; 99US-0152317.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ni J, Baker KP, Birse CE, Fiselja M, Komatsoulis GA, Rosen CA;
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Latifur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX WPI: 2001-203081/20.
DR P-PSDB; AAB87359.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
PS
XX Claim 1: Page 498; 607pp; English.
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC AAB87414-AAB87453 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
XX SQ Sequence 867 BP; 233 A; 180 C; 197 G; 256 T; 1 other;

Query Match 14.8%; Score 842; DB 22; Length 867;
Best Local Similarity 100.0%; Pred. No. 1.2e-221;
Matches 842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4848 ctggcattggccctgtgtctatcatgagggagagctaaagaagaattctctaggaag 4907
DB 13 ctgcattggccctgtgtctatcatcagagggagagctaaagaagaattctctaggaag 72
QY 4908 agctaatggccacagatcctcagtaattatattagtttggcttcgacagcttgt 4967
DB 73 agctaatggccacagatcctcagtaattatattagtttggcttcgacagcttgt 132

QY 4968 caggagggccacagaatgagacagataaaccagacagctcatcttgcctcctctac 5027
DB 133 caggagggccacagaatgagacagataaaccagacagctcatcttgcctcctctac 192
QY 5028 ggttttcaagtcagaggaattgtgtctgtctaatgtatgtatccaaactgtcaltgtga 5087
DB 193 ggttttcaagtcagaggaattgtgtctgtctaatgtatgtatccaaactgtcaltgtga 252
QY 5088 ctggcgatgccagctctccaaagtaattctgtcatgaaacataaagaagaagaagcc 5147
DB 253 ctggcgatgccagctctccaaagtaattctgtcatgaaacataaagaagaagaagcc 312
QY 5148 gttaacacatcacacagagaacatttcgggtcccaacagcgtgtgtgcaggaagctcact 5207
DB 313 gttaacacatcacacagagaacatttcgggtcccaacagcgtgtgtgcaggaagctcact 372
QY 5208 ctccgctcagatattgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5267
DB 373 ctccgctcagatattgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 432
QY 5268 ttcatgtttctgaacatcctgtatgttaacacatgctgggggtgctaaagtgcgtgaa 5327
DB 433 ttcatgtttctgaacatcctgtatgttaacacatgctgggggtgctaaagtgcgtgaa 492
QY 5328 tcccgatgtgaaanaagcgtgaggtgaagctcagacatcatgtattactttaanaac 5387
DB 493 tcccgatgtgaaanaagcgtgaggtgaagctcagacatcatgtattactttaanaac 552
QY 5388 aaaaaaaagacatgatatgtatgttttttttttttttttttttttttttttttttttttt 5447
DB 553 aaaaaaaagacatgatatgtatgttttttttttttttttttttttttttttttttttttt 612
QY 5448 gttagacttttttgaagaatgtatgttccacacacagctacccgtctcttcgtacattct 5507
DB 613 gttagacttttttgaagaatgtatgttccacacacagctacccgtctcttcgtacattct 672
QY 5508 gtgtcaatgtctcttcttcttaatacagctgcggcgtgtctaaagtgtgtttacacgtgtac 5567
DB 673 gtgtcaatgtctcttcttcttaatacagctgcggcgtgtctaaagtgtgtttacacgtgtac 732
QY 5568 ggcgagtgacctgtgtatgacgtgtcttcttcacagacctcccgagctgtgagaanaac 5627
DB 733 ggcgagtgacctgtgtatgacgtgtcttcttcacagacctcccgagctgtgagaanaac 792
QY 5628 agcttctctgtacatatacactccttaataaaggacatattctctcgtttaaaaaaa 5687
DB 793 agcttctctgtacatatacactccttaataaaggacatattctctcgtttaaaaaaa 852
QY 5688 aa 5689
DB 853 aa 854

RESULT 7
AAS27320
ID AAS27320 standard; cDNA; 823 BP.
XX
AC AAS27320;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 355.
XX
KW Neuroprotective; cytosolic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.

XX Homo sapiens.
OS
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
XX

PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239355.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249221.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
PR 05-JAN-2001; 2001US-0254097.
PR 05-JAN-2001; 2001US-0254097.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
DR MPI: 2001-465460/50.
DR P-PSDB: AAU17403.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT proposing disorders related to the proteins, including cancers, immune
PR disorders and neuronal disorders
XX
XX
PS Claim 1: SEQ ID NO 355; 880pp; English.

XX
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26796-AAS27850 represent novel signal transduction
XX pathway protein coding sequences and PCR primers of the invention.
XX

Query Match 14.4%; Score 819.4; DB 22: Length 823;
Best Local Similarity 99.6%; Pred. No. 2e-215;
Matches 820; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 773 cgggtttgagagatatacccaatttctctgtagatatagtgaaagcctgtagaagtc 832
DB 1 cggntttgagagatatacccaatttctctgtagatatagtgaaagcctgtagaagtc 60
QY 833 ccaagatgagagcctctggaatcagtagtgccttcaagctgtagagcgagaa 892
DB 61 ccaagatgagagcctctggaatcagtagtgccttcaagctgtagagcgagaa 120
QY 893 cccctgggttattagtaaaacgattggagaagtggttaaagctgtaacatgaaatctt 952
DB 121 cccctgggttattagtaaaacgattggagaagtggttaaagctgtaacatgaaatctt 180
QY 953 ttccgtgagaatgattgcatgtcagatataatgtagcgaaactctgaaataagaatttg 1012
DB 181 ttccgtgagaatgattgcatgtcagatataatgtagcgaaactctgaaataagaatttg 240
QY 1013 aacaagacaacatatgttcgcgaagcagtgtagacccatcatgtgttcacgttg 1072
DB 241 aacaagacaacatatgttcgcgaagcagtgtagacccatcatgtgttcacgttg 300
QY 1073 ttccgtgagaatgattgcatgtcagatataatgtagcgaaactctgaaataagaatttg 1132
DB 301 ttccgtgagaatgattgcatgtcagatataatgtagcgaaactctgaaataagaatttg 360
QY 1133 attcaagcgttttagccttgagcagcagatataatgtagcgaaagtgtagaagtgtag 1192
DB 361 attcaagcgttttagccttgagcagcagatataatgtagcgaaagtgtagaagtgtag 420
QY 1193 ggttcaacagtgtagagagaccccgagctgtaacaccccgctgtagaagtagacttc 1252
DB 421 ggttcaacagtgtagagagaccccgagctgtaacaccccgctgtagaagtagacttc 480

QY 1253 actcaagactactcatagcgcacacccctcgggaaaaacacacatccgtccagctcgg 1312
DB 481 actcaagactactcatagcgcacacccctcgggaaaaacacacatccgtccagctcgg 540
QY 1313 cactcaagaatgatttagtagcagctgtagcagtggttatacaccacaaaaatagga 1372
DB 541 cactcaagaatgatttagtagcagctgtagcagtggttatacaccacaaaaatagga 600
QY 1373 agaggttaataatccagcttaagaagaagtgtagaagtggttagatccagactactca 1432
DB 601 agaggttaataatccagcttaagaagaagtgtagaagtggttagatccagactactca 660
QY 1433 gagaatgtaacaataggtggtcagctcccaatctatgtgaaaaacatctcccgagggg 1492
DB 661 gagaatgtaacaataggtggtcagctcccaatctatgtgaaaaacatctcccgagggg 720
QY 1493 cggcattcaagtggtgcgagcttaaggcagaagacagactatagaggttaagttag 1552
DB 721 cggcattcaagtggtgcgagcttaaggcagaagacagactatagaggttaagttag 780
QY 1553 attagtgagcaaatcccaagaggaattgttcgctgttag 1595
DB 781 gtttagtgagcaaatcccaagaggaattgttcgctgttag 823

RESULT 8

AAH33594
ID AAH33594 standard; cDNA; 823 BP.
XX
AC AAH33594;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:650.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis: detection;
KW colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
PN MO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000MO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX MPI: 2001-235357/24.
XX P-PSDB: AAG4163.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT
PS Claim 1: Page 2675; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 823 BP; 249 A; 183 C; 201 G; 188 T; 2 other;

Query Match 14.4%; Score 819.4; DB 22; Length 823;
Best Local Similarity 99.6%; Pred. No. 2e-215;
Matches 820; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 773 cgggtttgagcatatcccaactttctctgatatgtaagctgtagaagctcc 832
DB 1 cggntttgagcatatcccaactttctctgatatgtaagctgtagaagctcc 60
QY 833 ccaacgtagagggcctctggaatccatgtagtccttcagtgctcgagcgagaga 892
DB 61 ccaacgtagagggcctctggaatccatgtagtccttcagtgctcgagcgagaga 120
QY 893 cccctgggtattagtaaaacgattggagaaggtgttaagctgtaactgtaaatctt 952
DB 121 cccctgggtattagtaaaacgattggagaaggtgttaagctgtaactgtaaatctt 180
QY 953 ttctgtagaatgattgcatgttcaagatlaattgtagcagcttcgaatagaagattg 1012
DB 181 ttctgtagaatgattgcatgttcaagatlaattgtagcagcttcgaatagaagattg 240
QY 1013 aacaagcaacaacatgatttcgccaagcagatgctacaccatcatttggtccatgtg 1072
DB 241 aacaagcaacaacatgatttcgccaagcagatgctacaccatcatttggtccatgtg 300
QY 1073 ttctgtagaatgattgcatgttcaagatlaattgtagcagcttcgaatagaagattg 1132
DB 301 ttctgtagaatgattgcatgttcaagatlaattgtagcagcttcgaatagaagattg 360
QY 1133 attcaagccgttttagccctgacagcagatatttgaacaagaggtgtgaacagtgcag 1192
DB 361 attcaagccgttttagccctgacagcagatatttgaacaagaggtgtgaacagtgcag 420
QY 1193 ggcttacaacggtgacagagacaccccgatgaacaccccgctgaagagatagactctc 1252
DB 421 ggcttacaacggtgacagagacaccccgatgaacaccccgctgaagagatagactctc 480
QY 1253 actcaagactacctatagcgacacccctcggaacacacacatccgctccagctcg 1312
DB 481 actcaagactacctatagcgacacccctcggaacacacacatccgctccagctcg 540
QY 1313 caccctcagaatgtattgtagactgttaagcaggtgttaaaccccaaaaaataggca 1372
DB 541 caccctcagaatgtattgtagactgttaagcaggtgttaaaccccaaaaaataggca 600
QY 1373 agaggtcttaataatccagcttaagaaggttacagaaggttggtatccgcatcttcca 1432
DB 601 agaggtcttaataatccagcttaagaaggttacagaaggttggtatccgcatcttcca 660
QY 1433 gagatgtacaatagtggtgttcagctccaatctatgtgtgaaacacatctcccccggggg 1492
DB 661 gagatgtacaatagtggtgttcagctccaatctatgtgtgaaacacatctcccccggggg 720
QY 1493 cggccattcagatgacgacttaagcagagagagacagactatagaggttaattgagtag 1552
DB 721 cggccattcagatgacgacttaagcagagagagacagactatagaggttaattgagtag 780
QY 1553 atttagtgggcaaatcccaagaggaagtgttctgctgtgtgag 1595
DB 781 gtttagtgggcaaatcccaagaggaagtgttctgctgtgtgag 823
```

RESULT 9

```
AAH07109
ID AAH07109 standard; cDNA; 556 BP.
XX
XX AAH07109;
AC
XX 26-JUN-2001 (first entry)
DT
XX Human cDNA clone (5'-primer) SEQ ID NO:3944.
DE
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
OS
XX Homo sapiens.
PN
XX EP1074617-A2.
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sueiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 1; SEQ ID 3944; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 556 BP; 154 A; 129 C; 144 G; 126 T; 3 other;
```

Query Match 9.7%; Score 549.8; DB 22; Length 556;
Best Local Similarity 99.1%; Pred. No. 4.3e-141;
Matches 551; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 374 ttgagttcacaccttgaagctcttcgagcaaatatgctcttcattgttcgacgaagtgtg 433
DB 1 ttgagttcacaccttgaagctcttcgagcaaatatgctcttcattgttcgacgaagtgtg 60
```

```
QY 434 acccagcttaattggcctctccactctctgtaagtagtaatttctctcgaagagc 493
    |||
Db 61 acccagcttaattggcctctccactctctgtaagtagtaatttctctcgaagagc 120
QY 494 cttcaaggaaaaatcccaacgcgtgttcaacaacagctgtcttccaaagcagaacatg 553
    |||
Db 121 cttcaaggaaaaatcccaacgcgtgttcaacaacagctgtcttccaaagcagaacatg 180
QY 554 ctggaggtccttaaaaacctgcgacgagaagaatgaaaaactacagaagcctccgcgg 613
    |||
Db 181 ctggaggtccttaaaaacctgcgacgagaagaatgaaaaactacagaagcctccgcgg 240
QY 614 atactagttaactggttcaacaattccagagagacaatgctcgtctctagtgcga 673
    |||
Db 241 atactagttaactggttcaacaattccagagagacaatgctcgtctctagtgcga 300
QY 674 gtcaaccaatgtgtggcgaagctgctgagaaacaagaagatgagatggagcagaag 733
    |||
Db 301 gtcaaccaatgtgtggcgaagctgctgagaaacaagaagatgagatggagcagaag 360
QY 734 aggataacagtcgtgtgaacctgtgtgacatgtgtgacacggtttcggagcataacca 793
    |||
Db 361 aggataacagtcgtgtgaacctgtgtgacatgtgtgacacggtttcggagcataacca 420
QY 794 actttctctgtagatgatagtgtaaaagctcgttagaagtcctccaaagatggagcctctg 853
    |||
Db 421 actttctctgtagatgatagtgtaaaagctcgttagaagtcctccaaagatggagcctctg 480
QY 854 gaatcatgtatgtccttccagtgcctcgagcgcggaacccctggggttaattagtaaac 913
    |||
Db 481 gaatcatgtatgtccttccagtgcctcgagcgcggaacccctggggttaattagtaaac 540
QY 914 gattggaagaaggtgg 929
    |||
Db 541 gattggaagaaggtgg 556

RESULT 10
AAH71480
ID AAH71480 standard; cDNA; 597 BP.
XX
AC AAH71480;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 2754.
XX
KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
XX
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000MO-US33312.
XX
PR 08-DEC-1999; 99US-0169681.
XX
PR 21-DEC-1999; 99US-0171350.
XX
PR 14-MAR-2000; 2000US-0189315.
XX
PR 12-MAY-2000; 2000US-0203791.
XX
PR 09-JUN-2000; 2000US-0210600.
XX
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI; 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
XX and for assessing and detecting compounds for treating the cancer -
```

```
PS - Claim 1: Page 566; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy.

SQ Sequence 597 BP; 180 A; 123 C; 156 G; 138 T; 0 other:

Query Match 9.1%; Score 516.2; DB 22; Length 597;
Best Local Similarity 91.0%; Pred. No. 8,6e-132;
Matches 576; Conservative 0; Mismatches 18; Indels 39; Gaps 1;

QY 1402 acagaaggtttggatccagcatcacttccagagatgtaacaatagttgctcagctcca 1461
    |||
Db 1 acagaaggtttggatccagcatcacttccagagatgtaacaatagttgctcagctcca 60
QY 1462 atctatgtgaaaaacatctcccccggggggggccatcaagatggcggacttaaggca 1521
    |||
Db 61 atctatgtgaaaaacatctcccccggggggggccatcaagatggcggacttaaggca 120
QY 1522 ggaacagacttataagatggaatggaatggaatggaatggaatggaatggaatggaat 1581
    |||
Db 121 ggaacagacttataagatggaatggaatggaatggaatggaatggaatggaatggaatggaat 180
QY 1582 gtttcgctgttggaagacccaagaatggaagaaactgttgaccccttctgtcttcgcag 1641
    |||
Db 181 gtttcgctgttggaagacccaagaatggaagaaactgttgaccccttctgtcttcgcag 240
QY 1642 gaagagcgttccacccaagggaactgaatgcagagccaagcccgatgcagattccaaaa 1701
    |||
Db 241 gaagagcgttccacccaagggaactgaatgcagagccaagcccgatgcagattccaaaa 266
QY 1702 gaaacgaagcagaagaatgagatattgttcttaacactgataggcaccaggaattctg 1761
    |||
Db 267 -----gaaacgaagaagaatgagatattgttcttaacactgataggcaccaggaattctg 321
QY 1762 acattgaagtcctcaacttaattgaltcagatcgtcagggccttgggtcgaatgtaaaagt 1821
    |||
Db 322 acattgaagtcctcaacttaattgaltcagatcgtcagggccttgggtcgaatgtaaaagt 381
QY 1822 aaccggtcaaaagagaacccacgagatttgggaattcttgcgaagtcattatataaggca 1881
    |||
Db 382 aaccggtcaaaagagaacccacgagatttgggaattcttgcgaagtcattatataaggca 441
QY 1882 ggaacagcatcataagaatggaagctcgggtggaatgatacaactgatacagtaaatgga 1941
    |||
Db 442 ggggcagcatcataagaatggaagctcgggtggaatgatacaactgatacagtaaatgga 501
QY 1942 gaatccctgttgggcaagacaacaaagatgcaatggaacccctaagaaggtctatgtct 2001
    |||
Db 502 gaatccctgttgggcaagacaacaaagatgcaatggaacccctaagaaggtctatgtct 561
QY 2002 actgaaggaataaagaaggaatgatacagctt 2034
    |||
Db 562 actgaaggaataaagaaggaatgatacagctt 594

RESULT 11
AAC79294
ID AAC79294 standard; cDNA; 540 BP.
XX
AC AAC79294;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human lung tumour-specific cDNA #247.
XX
```

KW Lung tumour protein; lung cancer; cytostatic; vaccine; ss.
 XX Homo sapiens.
 OS
 XX WO200060077-A2.
 PN
 XX 12-OCT-2000.
 PD
 XX 30-MAR-2000; 2000MO-US08560.
 PP
 XX 02-APR-1999; 99US-0285323.
 PR 09-AUG-1999; 99US-0370833.
 PR 30-DEC-1999; 99US-0476235.
 PR 03-MAR-2000; 2000US-0518809.
 XX
 PA (CORI-) CORIXA CORP.
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H;
 DR WPI; 2000-638466/61.
 XX
 PT Novel lung tumor polypeptides and polynucleotides, useful for
 PT detecting, monitoring or treating cancer, especially lung cancer -
 PS Claim 3; Page 226-227; 243pp; English.
 XX
 CC The present sequence is given in a specification relating to compounds
 CC for therapy and diagnosis of lung cancer. Polypeptides comprising at
 CC least an immunogenic part of a lung tumour protein are disclosed.
 CC The polypeptides are useful for inhibiting the development of cancer,
 CC especially lung cancer. Samples of T cells expressing the polypeptides
 CC may be used to inhibit the development of cancer. The polypeptides are
 CC also useful for detecting and monitoring the progression of cancer,
 CC especially lung cancer.
 CC
 SO Sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other;
 50

Query Match 8.5%; Score 482.4; DB 21; Length 540;
 Best Local Similarity 97.4%; Pred. No. 1,7e-122;
 Matches 535; Conservative 0; Mismatches 2; Indels 12; Gaps 4;
 QY 1990 aggtctatgtctactgaagcaataaagcaatgatacagctattgttgcaggaga 2049
 DB 1 aggtctatgtctactgaagcaataaagcaatgatacagctattgttgcaggaga 2049
 QY 2050 ataagcaagtcgaatgagctgaatcactcctggagcccccctggagctgagctccatt 2109
 DB 61 ataagcaagtcgaatgagctgaatcactcctggagcccccctggagctgagctccatt 2109
 QY 2110 gaaacagcttgatgataagagaagaaattccctccctctctacagtgagattgag 2169
 DB 121 gaaacagcttgatgataagagaagaaattccctccctctctacagtgagattgag 2169
 QY 2170 gggcttgaatgaatccagcagaagaatgctgctcctcagtagaataagtgagtgag 2229
 DB 181 gggcttgaatgaatccagcagaagaatgctgctcctcagtagaataagtgagtgag 2229
 QY 2230 aaatacagctgcccctcagtgataatgcccagaatacctgctcattatagaagat 2289
 DB 232 aaatacagctgcccctcagtgataatgcccagaatacctgctcattatagaagat 2289
 QY 2290 gaaagttgcaagtgctctcccaacatctctgcaagctcctctccagctcccatat 2349
 DB 292 gaaagttgcaagtgctctcccaacatctctgcaagctcctctccagctcccatat 2349
 QY 2350 gatgtggggttctgacggcagatgctgtaactggcagaagctcagatagatga 2409
 DB 352 gatgtggggttctgacggcagatgctgtaactggcagaagctcagatagatga 2409
 QY 2410 gccgacgctcttctgagtcacagatgctgacagctctgcttcttcaag-agaaggaatt 2468
 DB 412 gccgacgctcttctgagtcacagatgctgacagctctgcttcttcaag-agaaggaatt 471

QY 2469 tggagctcag-agtatgcagaaaaacgca-caagcaatttcagatgccagtcattg 2526
 DB 472 tggagctcagagatgctcagaaaaacgcaacacacgaatttcagatgccagtcattg 2526
 QY 2527 gatttcgctt 2535
 DB 532 gatttcgctt 540

RESULT 12
 AAD23370
 ID AAD23370 standard; cDNA; 540 BP.
 XX
 AC AAD23370;
 XX
 DT 26-FEB-2002 (first entry)
 DE Human lung tumour-specific SCC1-568 cDNA.
 XX
 KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
 KW antisense-therapy; vaccine; immune response; lung cancer; SCC1-568; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200172295-A2.
 PD 04-OCT-2001.
 PP 28-MAR-2001; 2001MO-US09991.
 PR 29-MAR-2000; 2000US-0538037.
 PR 05-JUN-2000; 2000US-0588937.
 PR 18-AUG-2000; 2000US-0640878.
 PR 22-SEP-2000; 2000US-234517P.
 PR 01-NOV-2000; 2000US-0704512.
 PR 14-DEC-2000; 2000US-0738973.
 PA (CORI-) CORIXA CORP.
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
 PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
 DR WPI; 2001-639201/73.
 XX
 PT New human lung-specific polynucleotides and polypeptides for the
 PT diagnosis and treatment of disease e.g. lung cancer -
 PS Example 2; Page 276; 378pp; English.
 XX
 CC The invention relates to isolated lung tumour-specific proteins and
 CC their corresponding cDNA molecules. Lung tumour-specific proteins and
 CC expanding T cells specific for a tumour protein, and for inhibiting
 CC the development of cancer. The invention also relates to a composition
 CC lung tumour specific oligonucleotide is useful in gene therapy and for
 CC diagnosis, detection and treatment of lung cancer. The present sequence
 CC is human lung tumour-specific cDNA.
 CC
 SO Sequence 540 BP; 147 A; 123 C; 133 G; 136 T; 1 other;
 50

Query Match 8.5%; Score 482.4; DB 23; Length 540;
 Best Local Similarity 97.4%; Pred. No. 1,7e-122;
 Matches 535; Conservative 0; Mismatches 2; Indels 12; Gaps 4;
 QY 1990 aggtctatgtctactgaagcaataaagcaatgatacagctattgttgcaggaga 2049
 DB 1 aggtctatgtctactgaagcaataaagcaatgatacagctattgttgcaggaga 2049
 QY 2050 ataagcaagtcgaatgagctgaatcactcctggagcccccctggagctgagctccatt 2109
 DB 61 ataagcaagtcgaatgagctgaatcactcctggagcccccctggagctgagctccatt 2109

XX DE Novel human polynucleotide, SEQ ID NO: 3345.
 XX KW Human; cytosolic; gene therapy; colon cancer; prostate cancer;
 XX KW breast cancer; lung cancer; cancer detection; ss.
 OS Homo sapiens.
 XX PN W0200102568-A2.
 XX PD 11-JAN-2001.
 XX PF 30-JUN-2000; 2000MO-US18374.
 XX PR 02-JUL-1999; 99US-0142310.
 XX PR 02-JUL-1999; 99US-0142311.
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
 PI Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 XX Kita D, Garcia V, Jones LM, Strache-Crain B;
 DR WPI: 2001-091805/10.
 XX PT Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -
 XX PS Claim 9; Page 1045; 1046pp; English.
 XX CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes of therapies and
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia, and
 CC dysplasia and hyperplasia.
 XX SQ Sequence 425 BP; 134 A; 86 C; 110 G; 95 T; 0 other;
 Query Match 73%; Score 415.4; DB 22; Length 425;
 Best Local Similarity 99.8%; Pred. No. 4,7e-104;
 Matches 416; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1732 cttaacactgtagcagcaggaattctgacattgaagtcacactaattgacaga 1791
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 9 cttaacactgtagcagcaggaattctgacattgaagtcacactaattgacaga 1791
 QY 1792 tctgagagcctgtgctcaggtcaaggtcaacgggtcaaaagagacacagcagattg 1851
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 69 tctgagagcctgtgctcaggtcaaggtcaacgggtcaaaagagacacagcagattg 1851
 QY 1852 ggaactctgtcagcagcattatattgagagagcagcagcctcaagatgnaagcttcg 1911
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 129 ggaactctgtcagcagcattatattgagagagcagcagcctcaagatgnaagcttcg 1911
 QY 1912 gtgatatcaactgtagcaggaattgagagatccctgttggcagagaacaacagagt 1971
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 189 gtgatatcaactgtagcaggaattgagagatccctgttggcagagaacaacagagt 1971
 QY 1972 gccatggaaccctaaagagcgtctatgtctactgaaggcaataaagaggaattccag 2031

DB 249 gccatggaaccctaaagagcgtctatgtctactgaaggcaataaagaggaattccag 308
 QY 2032 ctattgttgcagaggaataagcaagtcagagcgtgaagtcaccttggagccccct 308
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 309 ctattgttgcagaggaataagcaagtcagagcgtgaagtcaccttggagccccct 3091
 QY 2092 ggaacttgagcctgccatttgaacacagcgttggatgataagagcaaggaattcccat 2148
 DB 369 ggaacttgagcctgccatttgaacacagcgttggatgataagagcaaggaattcccat 425
 RESUOT 15
 ID AAH12111/c
 XX AAH12111 standard; cDNA; 564 BP.
 AC AAH12111;
 XX 26-JUN-2001 (first entry)
 DT XX
 DE Human CDNA clone (3'-primer) SEQ ID NO:8946.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 XX EP1074617-A2.
 PD 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 XX full-length cDNAs -
 XX PS Claim 3; SEQ ID 8946; 2537pp + CD ROM; English.
 XX CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any special methods. AAH03166 to AAH13628 and
 CC AAH95893 represent human amino acid sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.
XX
SQ Sequence 564 BP; 111 A; 141 C; 101 G; 200 T; 11 other;

Query Match		6.8%;	Score 384.4;	DB 22;	Length 564;
Best Local Similarity		90.6%;	Pred. No. 2.1e-95;		
Matches 415; Conservative		0;	Mismatches 41;	Indels	2; Gaps 1;
QY	2628 agatctgggtccctccctgggtctgaagaagcaagctcgcttgagagagctgcagacgcg	2687			
DB	564 AAAAATGGGGCCCTCCCGNGGNT--AAAAGCNAGCCCTTGAAAGTCTCCAGCCCC	507			
QY	2688 agctccgaggtgacttgaaatgggataatccctccalcgctccagcgccgataat	2747			
DB	506 AGTTNCCGAGGNCCTTTGAATGGNATATTCCTTTCATGGTCCACGGCCGCGATAAT	447			
QY	2748 cagaagcaggagatgcaatgagaagcttcagagctgcacatcgacaatctatgataacc	2807			
DB	446 CAGAGCGAGGGATGCAATGAGAGCTTCAGAGCTGCCATCGACAAATCTTATGATAACC	387			
QY	2808 cgcgttagatgatatgataaggaagcattgagaccttgagaagacacagaaagaaatc	2867			
DB	386 CGCGGTAGATGATGATGATGAAGGCATGAGACCTTGGAAGAGNCAGAAAGAAAGTTC	327			
QY	2868 aagatcaaggagagagagctgtatccacagccagtgatcagccttcccactctctgagag	2927			
DB	326 AAGATCAGCGAGAGAGTCTGTATCCACACCAGTGATCAGCCTTCCACTCTCTCGAGAG	267			
QY	2928 acaaatgaaatgaaaccaaagaagaagtgataagactgtatagaanaaagataaaactgg	2987			
DB	266 NCNAATGAATGGAACCAAGAGAAAGTGATAGACTGTAGAAAAAGGATAAAACTGG	207			
QY	2988 aaaagaaaaaagaagaagatagagataaggagaaggaataaatgaagaagccaagaaggaat	3047			
DB	206 AAAAGAAAAGAAAGAAAGATAGACATAGAGAAAGATAAATCAAAAGCCAAAGAGGGAAT	147			
QY	3048 gctgaagggtctggagagacatgltcaaglttgcaaac	3085			
DB	146 GCTGAAGGCTTGAGAGACATGTTCAGCCTTGCCAAAC	109			

Search completed: July 24, 2002, 11:20:04
Job time: 11721 sec

full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 18775; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dt primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 865 AA:

Query Match 62.0%; Score 4361; DB 22; Length 865;
Best Local Similarity 94.8%; Pred. No. 3,1e-301;
Matches 852; Conservative 5; Mismatches 8; Indels 34; Gaps 2;

QY 136 MFLHRRSSDPALIGLSTSVSDSNFSESRKRNPTMTAGFLKONTAGSPKTCDRKK 195
Db 1 mFLHRRSSDPALIGLSTSVSDSNFSESRKRNPTMTAGFLKONTAGSPKTCDRKK 60
QY 196 DENRSLPRDTSNMSNOFORNARSSLASHPMYKWLKQEDDEGTEDNSRVPVGH 255
Db 1 denrslprdtswnsnqfqrharslsashpmvkwkqeqdtdednsrvpvg 120
QY 256 ADTGLEHLPNFSLDWMKLVKVEVPNDGPGPLGHVVPSSARGSGRTGLVYKRLKGAKEH 315
Db 121 adtglehlpnfsldwmkvlvkvvevpndgpgplghvvpfsarggrtglvlykrlkgykakehe 180
QY 316 NLFRRNDQVIRINDGDLKRRREEQAKHMFROAMRTPIIFHVPDANKEOYEOLSOSEKN 375
Db 181 nlfrrndqvrrindgdlkrrrreeqakhmfroamrtpiifhvpdankkeqyeolseskn 240
QY 376 NYSSNFSPDSQYIDNRSVNSAGLFTVQAPRLNHPPOIDSHSLPSSAPSKPPSPAP 435
Db 241 nyssnfspdsqyidnrsvnsagltvqraprlnhppoidshslpssapksppspap 300
QY 436 ASAPQWVFSTVSSGNTFKIGKRLNIOLKGTGEGSITSRDVITGSAPIYKNTLP 495
Db 301 asapqvftstvsygnfkkikrlniqlkgtleglfsitardvtlgsaspiyknllp 360
QY 496 RGAIDQGRKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGCTSLVFEREDAFHPR 555
Db 361 rgaaidqgrkagdrllievngvdlvgksoeyvslrstkmgtvslvferedafhp 420
QY 556 LNAEPSONQIPKFKAKEDDVIYLPDGTREELFEVPLNDGSGAGLVSKNRSKENHA 615
Db 421 lnaepsonqipkfkakadedviltpdgtreelfevplndgsaglvsknrskenha 480
QY 616 DLGIFVKSIIINGAASKDGRILRVNDOLIAVNGSLLGKTNQDAMFRLRSMSREGKRCM 675
Db 481 dlgifvksiiingaaaskdgrilrvndolilavngsllgktnqdamefrrlsmsregkr 540
QY 676 IQLIVARRISKCNELKSPGPPPELPIETALDRERRISHLSYSGIEGLDESPSNAL 735

Db 541 IGLIVARRISKCNELKSPGPPPELPIETALDRERRISHLSYSGIEGLDESPSNAL 600
QY 736 SRIMGESGKYQLSPYVNMPODDTVIIIBDRLPYLPRLHSDQSSSHDDVGFVTFADAGTW 795
Db 601 srimgesgkyqlspyvnmppoddtviiibdrlpylprlhshdqssshddvgfvtadagtw 657
QY 796 AKAIISDACSISPDVDPVLAFOREGFGQSMSEKRTKOFSDASOLDPVTKRKSKMDL 855
Db 658 akaaisdacsispdvdpvylafoREGFGQSMSEKRTKOFSDASOLDPVTKRKSKMDL 689
QY 856 GIADETKINTVDDQKAGSPSRDVGPSGLKSSLSLSLQTAVAEYTLNGBIPFRPRPI 915
Db 690 geadetkintvddqkagspsrdivgpsglksslsleslqtavaevtnlgbipfrprpi 746
QY 916 IYRGCSNESFRAIDKSYDKPAVDDDDGMEETLEEDTEESSRSGRESVSTASDQPSHLE 975
Db 747 iyrqcsnesfraidksydkpavddddgmetleedteessrgresvstasdqpsshle 806
QY 976 ROMNGOEKGDTRDKKDKTKGKRRKRDKEKDKAKKGMKLGDMFRGRKRRKDX 1034
Db 807 romngoekgdtrdkdktkgkrrkrdkekdakgmklgdmfrgrkrrkdx 865

RESULT 2
AAG67610 standard; Protein; 865 AA.
ID AAG67610; XX

AA67610;

26-NOV-2001 (first entry)

Amino acid sequence of a human protein kinase/protein phosphatase.

Human; protein kinase; protein phosphatase; signal transduction.

Homo sapiens.

WO200109316-A1.

08-FEB-2001.

28-JUL-2000; 2000WO-JP05061.

29-JUL-1999; 99JP-0248036.

18-OCT-1999; 99US-0159590.

11-JAN-2000; 2000JP-0118776.

17-FEB-2000; 2000US-0183322.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
Senoo C, Nezu J;

WPI; 2001-570286/64.

N-PDB; AAH78369.

New genes encoding proteins with protein kinase/protein phosphatase activity, useful in the diagnosis and treatment of diseases -

Claim 2; Page 76-82; 233pp; Japanese.

The present sequence represents human protein kinase/protein phosphatase. It is expected that the protein kinase/protein phosphatase gene participates in signal transduction in cells. The protein kinase/protein phosphatase polypeptides and polynucleotides are useful for developing diagnostics and treatment agents for human and animal diseases. The protein kinase/protein phosphatase polypeptides are useful as target molecules in designing novel drugs. The protein kinase/protein phosphatase polynucleotides are useful as a source of probes and

CC primers, which may be used to isolate homologous sequences.

XX Sequence 865 AA:

Query Match 62.0%; Score 4361; DB 22; Length 865;
Best Local Similarity 94.8%; Pred. No. 3,1e-301;
Matches 852; Conservative 5; Mismatches 8; Indels 34; Gaps 2;

136 MCLHVRSSDPAALIGLSTVSQSNFSESPSRKRNPRMTAGLTKOMNAGSPKTDK 195
1 mclhvrssdpaliglstvsqsnfseesprknprrwtaltgilkqatgspkcdrrk 60
196 DENYRSLPDDTSNMWNOFQDNARSSLSASHPWGMKLEKQEDDEGTEEDNRPVGH 255
1 denyrslpddtsnmwnqfgrdnarslsashpmvgkqeqdeedgtesnrvpvyh 120
256 ADTGLEHIDNFSLDNWKLEVPPNDGPGIGIHVPFSARGRTLLGLVRLKRGKRAHE 315
1 adtglehpnfsldnmwklevppndgpglghvfpfsargrtllgllvkrlekrgkaeh 180
316 NLFREDCIVRIINDGDLRRRRFEQAQMFROAMRPTIIFHVVPANKRQYBQLSGSEKN 375
1 nlfredecivrlindgdlrrrrfeqaqhmfrqamrptli fhvvpankreqyqlsgsekn 240
376 NYSSRFSFSDQYIDNRSYNASAGLHTVQAPRLNHPPEOIDSRSRLPHSAHPGKPPSAP 435
1 nyssrfsfspdsgyidnrsynasaglhvqraprlnhppeqidsrslphsahpgkpsap 300
436 ASAPQWSTYVSSGYNTRYKIGKRLNIQLKGTGEGSITSRDVTIGSAPLYVKNILP 495
1 asapqwfstvyssgyntrykigkrlniqlkgtgegitsrdvtigsaplyvknllp 360
496 RGAIDODGRLLKAGDRILEVNGVDLVGKSOEYVSLRSTRMEGTSLVLRQDAPHPRE 555
1 rgaidodgrllkagdrilevngvdlvgkseeyvslrstremehtslvvrqdaahpre 420
556 LNAEPSQOMIPEKTKADEDIYLTDPGTREFLTAEVPLNDGSGAGLVGYSVKNRSEKNA 615
1 lnaepsqomipektaadediyldpghtrefltfevplndsgaglvgsyvknsrenha 480
616 DGIPTFKSIINGAASAKDGRRLRYNDOLIANVGSLLGKTQDMAMETLRKMSLEGKRM 675
1 dgiptfkssiingaaakdgrlryndqliavngesllgktqdametlrmslegnkrqm 540
676 IOLIVARRISKCELEKSPGSPPELPIETALDREKRISHLSYSGIEGDESRRNAL 735
1 iolivarriiskcelkspgsppeelpietaldrekrishlsysgiegdessrrnaal 600
736 SRIAGESGKYQLSPVTNMPDDTVIIEDDRPLVPLPHLSDQSSSSSHDDVGFYTADAGTW 795
1 sriagesgkyqlspvtnmppddtviieddrplvplphlsdosssshddvgfytadagtw 657
601 srim---gkyqlspvtnmppddtviieddrplvplphlsdosssshddvgfitaadgvt 657
796 AKAAISDASDCSLSPVDVPLAFQRBGFGQNSSEKRTKOFSDAQLDLYKTKRSKSM 855
1 akaaidsadcsldspvdvplafqrbgfgqnssekrkofsdasqlldlyktrksksmd 689
856 GIADERTKLVTDQKAGSPRDVPSLGKSSSLESLOTAVAEVTLNGDIPHRPRRI 915
1 giadertklvtdqkagspdvpslgkssssleslotavaevtlngdiphrprri 746
916 IGRGCGNESFRRAIDRSYDKPAVDDDEGMTLEEDTEESSRSGRESVSTASDQSHSL 975
1 igrsgcnesfraaidrsydkpavdddegmetleedteessrsgresvstasdqshsl 806
747 IIRGCGNESFRAIDRSYDKPAVDDDEGMTLEEDTEESSRSGRESVSTASDQSHSL 806
976 ROMNGNOEGDKTDKDKKDKTGKFKKKDKROKEMKRAKKAMLGIDMFEFGKHAKDK 1034
1 romngnoegdktdkdkdktgkfkdkrokemkrakkmllgidmfefgkhakdk 1034
807 RGMNGNQEGDKTDKDKKDKTGKFKKKDKROKEMKRAKKAMLGIDMFEFGKHAKDK 865
1 rgmngnqegdktdkdkdktgkfkdkrokemkrakkmllgidmfefgkhakdk 865

RESULT 3
AA771410
ID AA771410 standard; Protein: 744 AA.

XX AA771410;
AC 04-OCT-2000 (first entry)
DT
XX
DE PDZ domain-comprising PHIP (ephAin Interacting Protein) complex.
XX
KW PDZ domain: B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
KW cellular process; nerve cell interaction; regeneration of nerve cell;
KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;
KW differentiative disorder; PHIP; ephrin interacting protein; syntenin.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT Domain 273..351
FT Domain /label= PDZ_domain
FT Domain 461..546
FT Domain /label= PDZ_domain
FT Domain 590..678
FT Domain /label= PDZ_domain

MO200031124-A2.

02-JUN-2000.

19-NOV-1999; 99WO-CA01101.

20-NOV-1998; 98US-0109158.

(MOUN) MOUNT SINAI HOSPITAL.

Lln D, Pawson A;

WPI: 2000-400038/34.

N-PSDB: AAD01182.

Isolated complex for treating proliferative or differentiative disorders comprises B class ephrin and PDZ domain containing protein -

Claim 26; Fig 2D; 59pp; English.

The patent discloses a complex comprising of a B class ephrin and PDZ domain containing protein. B class ephrins function as ligands for Eph receptor tyrosine kinases (RTK) and possess a transmembrane element and a highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ binding site. This complex is used in methods to modulate the interaction of a B class ephrin and PDZ domain containing protein and to identify modulator compounds. It is also used for modulating cellular processes like, axonogenesis, nerve cell interactions and regeneration of nerve cells. The complex is also useful for treating proliferative or differentiative disorders associated with this protein complex. The present sequence is the PHIP (ephAin Interacting Protein) complex, comprising three PDZ domains. PHIP cDNA is isolated by screening the lambdaExioX 10.5 day mouse embryo expression library. PHIP is closely related to Caenorhabditis elegans PAR-3 protein, involved in regulating polarity of the early embryo. PHIP sequence can be used as a probe, to isolate other proteins like GRIP and syntenin.

Sequence 744 AA:

Query Match 50.1%; Score 3527; DB 21; Length 744;
Best Local Similarity 91.4%; Pred. No. 5,1e-242;
Matches 679; Conservative 26; Mismatches 38; Indels 0; Gaps 0;

1 MKYVFCGRTRVVVPCGDGHHKVFSLIQAVTRYRAIADPNWYIOVHRLHEDGILD 60
1 mkyvfcgrtrvvvpcgdgghkvfsliaqvtryraiadpnwyiovhrlhgdgild 60
61 LDOTLIDVADDKRLVAVPEODPHHGDCGTSASSTQSGEITGSGELGNNVAPRQY 120
1 ldotlidvadddkrlvavpeodphhgdgtsasstqsgelitgsgelgnnvaprqy 120
61 lddllcdvaddkrlvavfdeqdbphhgddgtsasstqsgelitgsgelgnnvaprqy 120

PE	17-JAN-2001	2001MO-US01312
XX	31-JAN-2000	2000US-1175065
PR	04-FEB-2000	2000US-0180658
PR	24-FEB-2000	2000US-0184664
PR	02-MAR-2000	2000US-0168550
PR	16-MAR-2000	2000US-0188974
PR	17-MAR-2000	2000US-0190076
PR	18-APR-2000	2000US-0196123
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209544
PR	28-JUN-2000	2000US-0214866
PR	30-JUN-2000	2000US-0215135
PR	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216680
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0225247
PR	14-AUG-2000	2000US-0225247
PR	14-AUG-2000	2000US-0225257
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	22-AUG-2000	2000US-0226821
PR	22-AUG-2000	2000US-0226868
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227109
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235583
PR	27-SEP-2000	2000US-0235832
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCT INC.
XX
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX MPI; 2001-465460/50.
XX
XX N-PSDB; AAS27320.
XX

PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PS disorders and neuronal disorders -
XX
XX
PS Claim 1: SEQ ID No 968; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), Ischaemic Injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Query Match 19.9%; Score 1397; DB 22; Length 273;
Best Local Similarity 99.3%; Pred. No. 3, 4e-91;
Matches 270; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 260 LEHINPESLDDMKYKLYEVNDGSPGILHYVPFARAGRTIGLYKLEKGRKHEHETLR 319
DB 2 LEHINPESLDDMKYKLYEVNDGSPGILHYVPFARAGRTIGLYKLEKGRKHEHETLR 61
QY 320 ENDCIYRINDGDLRNRFRPOAOHMFROAMRTPIITWVHPAANKEDYEQLOSEKNNYYS 379
DB 62 ENXCIVYRINDGDLRNRFRPOAOHMFROAMRTPIITWVHPAANKEDYEQLOSEKNNYYS 121
QY 380 SREPSDQYIDNRSVNSAGLHTVQARPLNHPPEQIDSHSRPLPHSAHPSGKPSAPASAP 439
DB 122 SREPSDQYIDNRSVNSAGLHTVQARPLNHPPEQIDSHSRPLPHSAHPSGKPSAPASAP 181
QY 440 QNWFSTTVSSGYTKKIKGRNLQILKKGTEGLFSITSRYVTGGSPRIYKNIILRGAA 499
DB 182 QNWFSTTVSSGYTKKIKGRNLQILKKGTEGLFSITSRYVTGGSPRIYKNIILRGAA 241
QY 500 IODGRKAGDRLIEVNGVPLVGSQSEWVSL 531
DB 242 IODGRKAGDRLIEVNGVPLVGSQSEWVSL 273

RESULT 5
AAG74163
ID AAG74163 standard; Protein; 272 AA.
XX
XX AAG74163;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:4927.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO200122920-A2.
XX
XX PD 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.
 PF
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA.
 DR WPI; 2001-235357/24.
 DR N-PSDB; AAH33594.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PS Claim 11; Page 6691-6693; 9803pp; English.
 XX
 CC AAH32943 to AAH32195 and AAH373514 to AAH37788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the activity of P by expressing
 CC additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SO Sequence 272 AA;

Query Match
 Best Local Similarity 19.8%; Score 1393; DB 22; Length 272;
 Matches 269; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 260 LEHITNESLIDMWKLVLEVPNDGPGPLHVVPSAAGCGRLGLVLRLEKCGKAHEHNEFR 319
 DB 2 LEHITNESLIDMWKLVLEVPNDGPGPLHVVPSAAGCGRLGLVLRLEKCGKAHEHNEFR 319
 QY 320 ENDCIVRINDGDLNRRFEQAOHMEFQAMRTPIHFHVVPANKEQYQOLSQSEKNYYS 379
 DB 62 ENDCIVRINDGDLNRRFEQAOHMEFQAMRTPIHFHVVPANKEQYQOLSQSEKNYYS 379
 QY 380 SRFSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSLRPHSAHPSGKPSAPASAP 439
 DB 122 SRFSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSLRPHSAHPSGKPSAPASAP 439
 QY 440 QNWFSTVSSGYNTKIKGRILNIOCKKTEGLGFSITSRDVTIGGSAPITYVKNILPRGAA 499
 DB 182 QNWFSTVSSGYNTKIKGRILNIOCKKTEGLGFSITSRDVTIGGSAPITYVKNILPRGAA 499
 QY 500 IDGGRKAGADRLIEVNGVDLVKSGQEEVSL 530
 DB 242 IDGGRKAGADRLIEVNGVDLVKSGQEEVSL 530
 DB 242 IDGGRKAGADRLIEVNGVDLVKSGQEEVSL 530

RESULT 6
 ABB61199
 ID ABB61199 standard; Protein; 1464 AA.
 XX AC ABB61199;
 XX
 DT 26-MAR-2002 (first entry)

XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 10389.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001MO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL05302.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 10389; 21pp + Sequence listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The development of
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB157737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIP0
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 1464 AA;

Query Match
 Best Local Similarity 16.4%; Score 1154.5; DB 22; Length 1464;
 Matches 433; Conservative 211; Mismatches 470; Indels 537; Gaps 61;

QY 1 MKTVCGGRTRVVPVPCGDSHMKVFSLIQOAVTRRKALADPNYVQVHRLHGDGDIID 60
 DB 1 MKTVCGGRTRVVPVPCGDSHMKVFSLIQOAVTRRKALADPNYVQVHRLHGDGDIID 60
 QY 61 LDDIICDVAHDRKRLVAVFEEDQ-----PHHGDDGTG-ASSTGOSPEFSELTGNNVS 114
 DB 59 LDDIICDVAHDRKRLVAVFEEDQ-----PHHGDDGTG-ASSTGOSPEFSELTGNNVS 114
 QY 115 APOPYQ-ATSELEWTPPS-----VLRANMPLHVRSSDPALIGSTVSNSFSSEPPSKN 169
 DB 117 PCPCPDSTPHLEVSTLSTGPMAGLGVLMVRRSSDPALIGSTVSNSFSSEPPSKN 169
 QY 170 PTKRSTTAGF-----LKONTAGSPKTCRKADEN-----YSLPDRDTSNMS 210
 DB 166 PTKRSTTAGF-----LKONTAGSPKTCRKADEN-----YSLPDRDTSNMS 210
 QY 211 NQFORNDARSSLSASHPMG-----KMLEKOE--ODEGTEE----- 245
 DB 225 NQFORNDARSSLSASHPMG-----KMLEKOE--ODEGTEE----- 245
 QY 246 -----DNSRVPVGHADTGLEHLPNFSLDNWKLVHVPNDGPGPLGIVHVPSPARGGRTL 299
 DB 285 -----DNSRVPVGHADTGLEHLPNFSLDNWKLVHVPNDGPGPLGIVHVPSPARGGRTL 299

Db	1276	ifnhrvghyaneydlhqqhqbqhlstrghyhsqraatzqdv-smhtstssgsgpslqap	1334
Oy	1189	ATOS-GRHSVS-----VEVOMQROBERESSQOAOQROXS.LP.ROSRKNAASVSDDSE-	1241
Db	1335	qagangryrrmasyyeyetvqgvtgskshhsasa-----tssssplnvphka	1384
Oy	1242	---ONYSPEGECPGSAKENPRYSSTYQGSRMNGYLCGHGFNARMVLEFQELLROE-----	1290
Db	1385	aamogyaspaslnssar-----srqftv-----lqvllregssgglpah	1422
Oy	1291	--QRRKEQOMKKOPP-----EGSNNDS	1312
Db	1423	llqghqgqqlqgqgqbtvqtkmspsqys	1453
RESULT	7		
ABG04715			
ID	ABG04715	standard; Protein; 169 AA.	
XX			
AC	ABG04715:		
XX			
DT	13-FEB-2002	(first entry)	
XX			
DE	Novel human diagnostic protein #4706.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
XX	food supplement; medical imaging; diagnostic; genetic disorder.		
OS	Homo sapiens.		
XX			
FN	WO200175067-A2.		
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
XX	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmamac RT, Liu C, Tang YT;		
DR	WPI: 2001-639362/73.		
DR	N-PSDB: AAS68902.		
XX			
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity		
PS	Claim 20; SEQ ID No 35074; 103bp; English.		
XX			
XX	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. ABG00010-ABG30377 represent novel human		
CC	diagnostic amino acid sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 169 AA:

Query Match 11.1%; Score 784; DB 22; Length 169;
 Best Local Similarity 93.8%; Pred. No. 6.5e-48;
 Matches 150; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1068 REFERRARREDAEAFIDFHTFGCDEELMYGVSYEGSMALNARPOSREGHMDALY 1127
 Db 9 REFERRARREDAEAFIDFHTFGCDEELMYGVSYEGSMALNARPOSREGHMDALY 68
 OY 1128 AOVKKPRNSKSPVDSNRSPSNHDIRIQRLRQEFQAKODEVDVDRRRRTYSFQOPWPNAR 1187
 Db 69 AQYKKPRNSKSPVDSNRSPSNHDIRIQRLRQEFQAKODEVDVDRRRRTYSFQOPWPNAR 128
 OY 1188 PATOSGRHSVSYEVOMQROREERESSOOAORXSSLPQ 1227
 Db 129 PATHSGRHSVSYEVQMHRQREERESSQAGQYSSLPQ 168

RESULT 8

ABG04716
 ID ABG04716 standard; Protein; 419 AA.

AC ABG04716;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #4707.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

XX N-PSDB; AAS68903.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX Claim 20; SEQ ID NO 35075; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 419 AA:

Query Match 8.8%; Score 616; DB 22; Length 419;
 Best Local Similarity 71.5%; Pred. No. 2.2e-35;
 Matches 123; Conservative 4; Mismatches 17; Indels 28; Gaps 1;

OY 1026 FCKHRRDKIEKTKIKIOESFTSEERIRNKQOEFLQAKTREFRQAREDAEAFID 1085
 Db 113 FGKHKKDKIEKTKIKIQESFTSEERIRNKQOEFLQAKTREFRQAREDAEAFID 172
 OY 1086 FHTFGCDEELMYGVSYEGSMALNARPOSREGHMDALYAOVKKPRNSKSPVDSNR 1145
 Db 173 FHTFGCDEELMYGVSYEGSMALNARPOSREGHMDALYAOVKKPRNSKSPVDSK 232
 OY 1146 STPSNHDRIQRLRQEFQAKODEVDVDRRRRTYSFQOPWPNARPATOSGRHSV 1197
 Db 233 KVLGEADPV-----YPRNPTEGNASL 256

RESULT 9

ABB1149
 ID ABB1149 standard; peptide; 315 AA.

AC ABB1149;

DT 11-JAN-2002 (first entry)

DE Human PAR3 homologue, SEQ ID NO:1519.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumor; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteopathic; vasotropic; cardiatic; virucide; antibacterial;
 KW antifungal; vulnery; antitumor.

XX Homo sapiens.

OS WO200157188-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US03800.

XX 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-457740/49.

XX N-PSDB; ABA08393.

PT Human proteins and DNA encoding sequences useful for preventing,

RESULT 11

AA71424

ID AAY71424 standard; Protein; 90 AA.

AC AAY71424;

DT 04-OCT-2000 (first entry)

DE PDZ domain-comprising PHIP PDZ3 protein complex.

KW PDZ domain: B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
 KW cellular process; nerve cell interaction; regeneration of nerve cell;
 KW axonogenesis; antiproliferative; proliferative disorder; treatment; GRIP;
 KW differentiative disorder; PHIP PDZ3 complex; ephrin interacting protein.

OS Mus sp.

PN WO20031124-A2.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-CA01101.

PR 20-NOV-1998; 98US-0109158.

PA (MOUN) MOUNT SINAI HOSPITAL.

PI Lin D, Pawson A;

DR WPI; 2000-400038/34.

PT Isolated complex for treating proliferative or differentiative
 disorders comprises B class ephrin and PDZ domain containing protein -

PS Claim 3; Fig 2C; 59pp; English.

CC The patent discloses a complex comprising of a B class ephrin and PDZ
 CC domain containing protein. B class ephrins function as ligands for Eph
 CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
 CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
 CC binding site. This complex is used in methods to modulate the interaction
 CC of a B class ephrin and PDZ domain containing protein and to identify
 CC modulator compounds. It is also used for modulating cellular processes
 CC like, axonogenesis, nerve cell interactions and regeneration of nerve
 CC cells. The complex is also useful for treating proliferative or
 CC differentiative disorders associated with this protein complex.
 CC The present sequence is the PDZ domain-comprising PHIP (ephrin
 CC interacting protein) PDZ3 protein complex. PHIP consists of three PDZ
 CC domains and is closely related to Caenorhabditis elegans PAR-3 protein,
 CC involved in regulating polarity of the early embryo. This sequence
 CC is isolated by screening the lambdaEX10x 10.5 day mouse embryo
 CC expression library.

SQ Sequence 90 AA;

Query Match Best Local Similarity 6.0%; Score 419.5; DB 21; Length 90;
 Matches 86; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

OY 590 EVP-LNDSGSAGLGVSQVKNRSEKNEHDLGIFPKSIINGGAASKDGLRYNDOLIVNGE 648
 DB 1 EEPIDSGSAGLGVSQVKNRSEKNEHDLGIFPKSIINGGAASKDGLRYNDOLIVNGE 60
 OY 649 SLUGKTQDAMETLRNSMSTEGNKKRGMQL 678
 DB 61 SLIGKANGAEMLTRSMSTEGNKKRGMQL 90

RESULT 12

AA71423

ID AAY71423 standard; Protein; 86 AA.

XX

AC AAY71423;

DT 04-OCT-2000 (first entry)

DE PDZ domain-comprising PHIP PDZ2 protein complex.

KW PDZ domain: B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
 KW cellular process; nerve cell interaction; regeneration of nerve cell;
 KW axonogenesis; antiproliferative; proliferative disorder; treatment;
 KW differentiative disorder; PHIP PDZ2 complex; ephrin interacting protein.

OS Mus sp.

PN WO20031124-A2.

PD 02-JUN-2000.

PF 19-NOV-1999; 99WO-CA01101.

PR 20-NOV-1998; 98US-0109158.

PA (MOUN) MOUNT SINAI HOSPITAL.

PI Lin D, Pawson A;

DR WPI; 2000-400038/34.

PT Isolated complex for treating proliferative or differentiative
 disorders comprises B class ephrin and PDZ domain containing protein -

PS Claim 3; Fig 2C; 59pp; English.

CC The patent discloses a complex comprising of a B class ephrin and PDZ
 CC domain containing protein. B class ephrins function as ligands for Eph
 CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
 CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
 CC binding site. This complex is used in methods to modulate the interaction
 CC of a B class ephrin and PDZ domain containing protein and to identify
 CC modulator compounds. It is also used for modulating cellular processes
 CC like, axonogenesis, nerve cell interactions and regeneration of nerve
 CC cells. The complex is also useful for treating proliferative or
 CC differentiative disorders associated with this protein complex.
 CC The present sequence is the PDZ domain-comprising PHIP (ephrin
 CC interacting protein) PDZ2 protein complex. PHIP consists of three PDZ
 CC domains and is closely related to Caenorhabditis elegans PAR-3 protein,
 CC involved in regulating polarity of the early embryo. This sequence
 CC is isolated by screening the lambdaEX10x 10.5 day mouse embryo
 CC expression library.

SQ Sequence 86 AA;

Query Match Best Local Similarity 5.8%; Score 410; DB 21; Length 86;
 Matches 84; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 461 NIQLKKGTEGLGFSITSRDVTIGSGAPIVYKNILPPGAAIOGRKAKGDRILEVNGVDLV 520
 DB 1 nqllkkgteglgfsitstrdvtlgsaaplyvknllprgaaiqgrlkaqurllvngvdl 60
 OY 521 GKSQEEVSLNKRSTKMEGVSLVFR 546
 DB 61 gksqeevslnkrstkmegvslvfr 86

RESULT 13

AAB01383

ID AAB01383 standard; Protein; 856 AA.

AC AAB01383;

DT 20-OCT-2000 (first entry)

XX	Neuron-associated protein.
DE	
XX	
KW	Neuron associated protein; NEUP; neurological disorder; epilepsy;
KW	ischemic cerebrovascular disease; stroke; cerebral neoplasm;
KW	Alzheimer's disease; Pick's disease; Huntington's disease;
KW	dementia; Parkinson's disease; demyelinating disease; meningitis;
KW	prion disease; Kuru; Creutzfeldt-Jakob disease; neurofibromatosis;
KW	cerebral palsy; muscular dystrophy; central nervous system; CNS;
KW	peripheral nervous system; PNS; myopathy; schizophrenia;
KW	acinic keratosis; arteriosclerosis; atherosclerosis; burstsitis;
KW	cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW	myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
KW	autoimmune disease; inflammation; acquired immunodeficiency syndrome;
KW	AIDS; Addison's disease; adult respiratory distress syndrome;
KW	allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
KW	Werner syndrome; trauma; human.
OS	
XX	
XX	homo sapiens.
XX	
PH	Key
PH	Location/Qualifiers
FT	1..65
FT	/label- PDZ domain signature
FT	/label- PDZ domain signature
FT	18
FT	/note- "Potential glycosylation site"
FT	20
FT	/note- "Potential phosphorylation site"
FT	69
FT	/note- "Potential phosphorylation site"
FT	71
FT	/note- "Potential phosphorylation site"
FT	73
FT	/note- "Potential phosphorylation site"
FT	118
FT	/note- "Potential phosphorylation site"
FT	119
FT	/note- "Potential phosphorylation site"
FT	136..218
FT	/label- PDZ domain signature
FT	144..147
FT	/note- "Glycosaminoglycan attachment site"
FT	199
FT	/note- "Potential glycosylation site"
FT	238
FT	/note- "Potential phosphorylation site"
FT	253
FT	/note- "Potential phosphorylation site"
FT	263
FT	/note- "Potential phosphorylation site"
FT	/note- "Potential phosphorylation site"
FT	269..349
FT	/label- PDZ domain signature
FT	284
FT	/note- "Potential phosphorylation site"
FT	296
FT	/note- "Potential phosphorylation site"
FT	337
FT	/note- "Potential phosphorylation site"
FT	341
FT	/note- "Potential phosphorylation site"
FT	369
FT	/note- "Potential glycosylation site"
FT	370
FT	/note- "Potential phosphorylation site"
FT	386
FT	/note- "Potential phosphorylation site"
FT	389
FT	/note- "Potential glycosylation site"
FT	390
FT	/note- "Potential phosphorylation site"
FT	414
FT	/note- "Potential phosphorylation site"
FT	/note- "Potential phosphorylation site"
FT	415..497
FT	/label- PDZ domain signature

FT	Modified-site	475	/note= "Potential phosphorylation site"
FT		511...592	
FT	Domain	/label= PDZ domain signature	
FT	Modified-site	531	/note= "Potential glycosylation site"
FT	Modified-site	568	/note= "Potential glycosylation site"
FT	Modified-site	599	/note= "Potential glycosylation site"
FT	Modified-site	604	/note= "Potential phosphorylation site"
FT	Modified-site	625	/note= "Potential phosphorylation site"
FT	Modified-site	633	/note= "Potential phosphorylation site"
FT	Modified-site	634	/note= "Potential phosphorylation site"
FT	Domain	648...733	/label= PDZ domain signature
FT	Modified-site	690	/note= "Potential phosphorylation site"
FT	Modified-site	705	/note= "Potential phosphorylation site"
FT	Modified-site	721	/note= "Potential glycosylation site"
FT	Domain	773...856	/label= PDZ domain signature
FT	Region	819...821	/label= Cell attachment sequence
FT	Modified-site	835	/note= "Potential phosphorylation site"
PN	WO20034477-A2.		
PD	15-JUN-2000.		
XX	10-DEC-1999;	99WO-US30408.	
XX	11-DEC-1998;	98US-0210083.	
PR	09-FEB-1999;	98US-0123456.	
PR	16-MAR-1999;	99US-0124687.	
PA	(INCY-) INCYTE PHARM INC.		
PI	Tang YT, Yue H, Baughn MR, Hillman JL, Lai P, Au-young J, Yang J		
PI	Lu DAM, Azimtal Y;		
XX	WPI: 2000-423423/36.		
DR	N-PSDB: AAA47424.		
XX	New human neuron-associated proteins and polynucleotides encoding them		
PT	useful for diagnosis, treatment and prevention of cell proliferative		
PT	disorders including cancer, neuronal and neurological disorders		
XX	Claim 1: page 105-108; 145pp; English.		
XX	Human neuron-associated proteins (NEUP) can be used for for		
CC	treating or preventing a disorder associated with decreased		
CC	expression or activity of NEUP. Antagonists of NEUP are useful for		
CC	treating or preventing disorder associated with increased expression		
CC	or activity of NEUP. NEUP or their fragments or derivatives are		
CC	useful for treating neurological disorder such as epilepsy, ischemic		
CC	cerebrovascular disease, Huntington's disease, dementia and		
CC	Parkinson's disease. NEUPs are also useful for treating other		
CC	demyelinating diseases, bacterial and viral meningitis, prion		
CC	diseases including kuru, Creutzfeldt-Jakob disease, nutritional and		
CC	metabolic diseases of the nervous system, neurofibromatosis, other		
CC	developmental disorders of the central nervous system, cerebral		
CC	palsy, neuroskeletal disorders, autonomic nervous system disorders,		
CC	cranial nerve disorders, spinal cord diseases, muscular dystrophy and		

CC other neuromuscular disorders, peripheral nervous system disorders,
 CC inherited, metabolic, endocrine, and toxic myopathies, mental
 CC disorders including mood, anxiety and schizophrenia disorders, a cell
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
 CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
 CC disease, adult respiratory distress syndrome, allergies, ankylosing
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
 CC complications of cancer, hemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic
 CC infections, and trauma. This protein was given the Incey ID no.
 CC 3201753CD1.
 CC
 XX Sequence 856 AA:
 SQ

Query Match 5.1%; Score 355.5; DB 21; Length 856;
 Best Local Similarity 26.0%; Pred. No. 2.1e-16;
 Matches 132; Conservative 69; Mismatches 187; Indels 119; Gaps 17;

OY 247 NSRVEPVGHADTGLEHINFSLDMMKLVENPDGPGIHVVPFSGRGRLGLVRL 306
 DB 245 nkeptvttsdaavdlstfk---nvghlelpkdgqglgla!-----seedtsgvllksl 297
 OY 307 EKGKAHEHNLFRENDCIVRINDGDLRRNRFEQAOHMFROAMRPIIMFH-----VVP 359
 DB 298 tehgvaaatdgrlkvgdqlavddelivgyprkfsltkakmtvklthlaenpdsqavp 357
 OY 360 AANKQEYEOLOSSEKNNYSSRFSPOSOYIDNRSVNSAGLHVQAPRLNHPROIDSHS 419
 DB 358 sa-----agaagekknssqslmvpqsg-----spepslntns 391
 OY 420 RLPHSAHPBGRP---PSAPASAPQNVFSTVSSGYNFKKIGKRLNIOAKTGGLGFSIT 476
 DB 392 r-----stspafasdpacp-----lpgcet-----tlelskgrtgyllslv 430
 OY 477 SRQVYIGGSAPLYVKNILPRGAIIDGRLKAGDRLILEVNGVDLVGKSOEVSILRSTKM 536
 DB 431 ggsdtllga---llheveyeegaackdgrlwgddqllvengldlrkathdeanlvrlqcpq 488
 OY 537 EGTVSLVLFROEDAFHRELNAEPQOMQIPRETKAEDDEDIVLPDGTREFLFEVPLNDS 596
 DB 489 r---vrltlyrdeapy-----keeevcd-----tltlelqkk 517
 OY 597 GSAGLGVSVKGNRSKENHADLGIIFVKSIIINGGAASKDGLRVLNDOLAVNGESILGKTNO 656
 DB 518 psgkglgslivgkn---dtgyfvsdvlvkgdladagrlmgdqlimvngedvratqge 572
 OY 657 DAMETLRRSMSTEGNKGMIQ---LIVARISKCNELKSPSGPPPELPETALDDNERR 713
 DB 573 aavaallkcsigtvtlevgrtkagprfhsetrrpsqsgv-segslasfcfplsgs----- 624
 OY 714 ISHSLVSGIEGIDESPRAALSRING 740
 DB 625 -----steelslesskknalaseiqg 645

RESULT 14
 AAY53753
 ID AAY53753 standard; Protein: 2037 AA.
 XX
 AC AAY53753;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the MMS2 protein.
 XX
 KW Human: MMS2; MMAC1; PDZ domain; tumour suppressor; tyrosine phosphatase;
 XX
 KW Scatolding protein; cancer.

OS Homo sapiens.
 XX
 PN W09958548-A1.
 XX
 PD 18-NOV-1999.
 XX
 PF 07-MAY-1999; 99WO-US09969.
 XX
 PR 08-MAY-1998; 98US-0084740.
 XX
 PA (MRT-) MYRIAD GENETICS INC.
 XX
 PI Bartel PL, Tavligian SV;
 XX
 DR WPI; 2000-053077/04.
 DR N-PADB; AA36453.
 XX
 PT Nucleic acids and polypeptides representing human MMS2, useful for
 PT detecting, diagnosing a predisposition to, and treating cancer -
 PS Claim 1; Page 93-99; 112pp; English.

The present sequence represents human MMS2 protein. The MMAC1 protein binds to MMS2. The MMS2 protein has 11 post-synaptic density protein, disc-large, zo-1 (PDZ) domains and one or more of these domains interact specifically with the carboxyl terminal amino acids of MMAC1 (see AAY53753). Specifically, it appears that domain 7, 10 and 13 interact with MMAC1. Since MMS2 contains 11 PDZ domains and interacts with MMAC1, a known tumour suppressor having a region of homology with common tyrosine phosphatases, MMS2 acts as a scaffolding protein in a interaction between MMAC1 and MMS2. It is believed that the activity of MMAC1. The MMS2 polypeptides, polynucleotides, fragments and or a predisposition to cancer and screening for agents that may be used to treat MMS2 and/or MMAC1 related cancer. The polypeptides and polynucleotides may also be used to treat cancer.

Sequence 2037 AA:

Query Match 5.1%; Score 355.5; DB 21; Length 2037;
 Best Local Similarity 26.0%; Pred. No. 7.6e-16;
 Matches 132; Conservative 69; Mismatches 187; Indels 119; Gaps 17;

OY 247 NSRVEPVGHADTGLEHINFSLDMMKLVENPDGPGIHVVPFSGRGRLGLVRL 306
 DB 1426 nkeptvttsdaavdlstfk---nvghlelpkdgqglgla!-----seedtsgvllksl 1478
 OY 307 EKGKAHEHNLFRENDCIVRINDGDLRRNRFEQAOHMFROAMRPIIMFH-----VVP 359
 DB 1479 tehgvaaatdgrlkvgdqlavddelivgyprkfsltkakmtvklthlaenpdsqavp 1538
 OY 360 AANKQEYEOLOSSEKNNYSSRFSPOSOYIDNRSVNSAGLHVQAPRLNHPROIDSHS 419
 DB 1539 sa-----agaagekknssqslmvpqsg-----spepslntns 1572
 OY 420 RLPHSAHPBGRP---PSAPASAPQNVFSTVSSGYNFKKIGKRLNIOAKTGGLGFSIT 476
 DB 1573 r-----stspafasdpacp-----lpgcet-----tlelskgrtgyllslv 1611
 OY 477 SRQVYIGGSAPLYVKNILPRGAIIDGRLKAGDRLILEVNGVDLVGKSOEVSILRSTKM 536
 DB 1612 ggsdtllga---llheveyeegaackdgrlwgddqllvengldlrkathdeanlvrlqcpq 1669
 OY 537 EGTVSLVLFROEDAFHRELNAEPQOMQIPRETKAEDDEDIVLPDGTREFLFEVPLNDS 596
 DB 1670 r---vrltlyrdeapy-----keeevcd-----tltlelqkk 1698
 OY 597 GSAGLGVSVKGNRSKENHADLGIIFVKSIIINGGAASKDGLRVLNDOLAVNGESILGKTNO 656
 DB 1699 psgkglgslivgkn---dtgyfvsdvlvkgdladagrlmgdqlimvngedvratqge 1753

```
OY 657 DAMELTRSMSTEGNKRGMIO---LIVARRISCKNELKSPGPPPELPIETALDDRRER 713
DB 1754 aavaalkcslgltvleqvgrfkagpfhserpsqagv-segslsftfpisg----- 1805
OY 714 ISHSLYSGIEGLDESPSRNALSLRING 740
DB 1806 -----stseslesskknalaseig 1826

RESULT 15
AAV04733
ID AAV04733 standard: Protein; 2070 AA.
XX
AC AAV04733;
XX
DT 06-JUL-1999 (first entry)
XX
DE Protein containing PDZ domain from clone 38-2-1b.
XX
KW PDZ domain; gene expression; human umbilical vascular endothelial cell;
KW HUVEC; stimulation; tumour necrosis factor; TNF; protein binding;
KW cell; proliferation disorder; cancer.
XX
OS Homo sapiens.
XX
PN WO9907046-A1.
XX
PD 18-FEB-1999.
XX
PF 12-AUG-1998; 98WO-JP03603.
XX
PR 19-JUN-1998; 98JP-0189944.
XX
PR 12-AUG-1997; 97JP-0230356.
XX
PA (CHUGAI) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Funahashi S, Miyata S;
XX
DR WPI: 1999-167423/14.
DR N-PSDB: AAX29910.
XX
PT Protein containing PDZ domain, whose expression is enhanced by TNF
PT stimulation - plays an important role in protein/protein
PT interactions and is used for screening for proteins for use in
PT treatment of cell proliferation disorders such as cancer
XX
PS Claim 1: Page 167-176; 240pp; Japanese.
XX
CC This sequence represents a new protein containing a PDZ domain encoded
CC by the clone 38-2-1b, whose expression in human umbilical vascular
CC endothelial cells (HUVEC) is enhanced by stimulation with tumour necrosis
CC factor (TNF) alpha. The new protein is used to identify proteins which
CC bind to it (particularly to the PDZ domains) and the genes encoding them,
CC for use in the treatment of cell proliferation disorders such as cancer.
XX
SQ Sequence 2070 AA:

Query Match 5.1%; Score 355.5; DB 20; Length 2070;
Best Local Similarity 22.5%; Pred. No. 7,8e-16;
Matches 188; Conservative 122; Mismatches 302; Indels 225; Gaps 37;

OY 230 GKWLKQEDDGDGTEDNSRVEPVGHADTGLEHHPNFSLDDNWKLYEVPNDGPGIGIHVV 289
DB 1119 grdlpelperseg-egeseelqntays-----nwnqprvelwrepeks--lglslv 1167
OY 290 PPSARCGRL-----GLVYKRLKGGKGAHEHNLFRENDCYVRINDGDLNRNRF 338
DB 1168 -----ggrgmgsrlsngewmrglflkhvledrpgkngtlkpgdrivewgmldirdashe 1222
OY 339 OAOHMFROAMRPIIWFHVVPANKEOYEQLSQSEKNNYSSRFPDSOYIDNRSVNSAG 398
DB 1223 qaveairka-gnpv-fmwgslnrprksplslmlhlypkynfstunpfadsiqina-- 1278
```

```
OY 399 LHTVORAPRLNHPPEQIDSHSLRPLPHSAHPGCKPPSAPAPQNVFTYVSS----- 449
DB 1279 -----dkapsqese---pekaplcsvpppppsafemsgsdhtgsaeklsqvd 1325
OY 450 -----GYNFKIKGRLN-----IQLKKTGCLGSGIT-SRDVYTGSGAPYIVKILP 495
DB 1326 kedefgyswknirerygtltgelhmlelekyhsglslagnkd---rsmvflvygldp 1382
OY 496 RGAALDGRKAGDRLELVNGVDLVGKSGEEVSLRSTKMEGVSLLVPRQEDAFPRE 555
DB 1383 ngaagkdgtrlgtdellelhnqglllytrshgnassllkcapsk--vklifrnkdavngma 1440
OY 556 L---NA-EPSOMQIPKTKADEDDIVLTPDGTREPLF-----EVLNDSGSAGLCVSV 605
DB 1441 vcpгнаveprpsnsenlnketeptltsdaavdlasfkvnqhlelp-kdgqglglals 1499
OY 606 KGNRSKRNHADLGIFVASTINGGAASKDGRLRVNDQILVANGESLQ-----KT- 654
DB 1500 edtlsl-----gvllkslshgvaaatdgrllkvydqllavddevlygplekflslkta 1552
OY 655 -----NDAMELTRSMSTEGNKRGMIOIIVA-----RISCKNELKS 692
DB 1553 kmrvklitihaeinpqsqavpsaagaasgekknssqslmvpqsgspsespesintrstcpa 1612
OY 693 PGSPPG--PELP-IEFALDDRRER-----ISHSLYSGIEGLDESPSR 731
DB 1613 faadpatcpllpccettelskgrtglslvgstdllgafllheve---egaackdgr 1670
OY 732 NAALSRIKESG---KYQLSPYVM---PQDDTVIIEDRLPLVPLPHLSQ----- 776
DB 1671 lwegdqlllevngldlrkathdeainvrtcpqrvtllyrdeapgykeeevcdltlelqk 1730
OY 777 -----SSSSSHDDVGFVTADAGTWAKAASPSADCSLSP-----VDP 814
DB 1731 kpgkglglslvgkrndtgvtrvsd---vkgyiaqpdgrlllgdqlllvngedvtnasgea 1787
OY 815 VLAFQREGFGROSN-----SEKR---TKQFSDAQQLDF-----KTRKSK 851
DB 1788 vaallkcsllgtvllevgrlkgpfhserprsqtsqvsagslsftfpisgstsless 1847
OY 852 SMDLGTIDENK-LNTVDDOKAGSPSRDVPGLGKSSLSLEQTANAETVINDIP 907
DB 1848 skknalaseigrltvemkk--gpldslgislagvgvsl-----gdvp 1889
```

Search completed: July 24, 2002, 14:29:50
Job time: 11382 sec

QY 597 GSAGLVSVKNSKENHADLGFVKSIIINGGAASKDGLRVNDQLIAVNGESILGKTNO 656
 Db 1699 PGKGLGSIYKRR-----DTGVFSDIYKGIADADRLKMGDQIILWNGEDVRNNTQ 1753
 QY 657 DAMETLRSMSTGKNGKMGIO---LIVARRISKCNELKSPSPGPPELPIETALDDERR 713
 Db 1754 AVAALLKCSIGVTTLEVGVKIKAGPFPHSRRPSQSV-SESLISFTFPLSGS----- 1805
 QY 714 ISHSLSYSGIEGLDESPSRNAALSRIING 740
 Db 1806 -----STSESLSSSKKNALASLEIIG 1826

RESULT 2
 US-09-233-086-3
 ; Sequence 3, Application US/09233086
 ; Patent No. 6337192
 ; GENERAL INFORMATION:
 ; APPLICANT: Bartel, Paul L.
 ; APPLICANT: Tavligian, Sean V.
 ; TITLE OF INVENTION: MMSCL- An MMSCL Interacting Protein
 ; FILE REFERENCE: MMSCL- An MMSCL Interacting Protein
 ; CURRENT APPLICATION NUMBER: US/09/233,086
 ; EARLIER FILING DATE: 1999-01-19
 ; EARLIER APPLICATION NUMBER: US 60/071,861
 ; NUMBER OF SEQ ID NOS: 65
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1881
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-233-086-3

Query Match 4.8%; Score 340.5; DB 4; Length 1881;
 Best Local Similarity 21.0%; Pred. No. 2.3e-16;
 Matches 185; Conservative 102; Mismatches 266; Indels 327; Gaps 35;

QY 124 ELEVPVSLRANMPLHVRSSDPALIGLSTVSDFSSE-----EPRSKNTMTSTA- 177
 Db 877 ELVDPSPSMELVPLSHIQEATPV-----PSVNEHLFGTQWLHNDPESQPARIGRTVY 931
 QY 178 -----GFLKONTAGSPTCKDKD-----ENTRSLPRDTSWMSNOFQ 214
 Db 932 SQEAPGYGCPENVMKENVFMELESPVSTEGNSOQGFEDLEINLSIAK-TSLDLGMP 990
 QY 215 RDNARSLASHPMVKMKLEKQDQEDGTEEDNSRVEPVGHADTGL----- 260
 Db 991 NDVQSPSLIDLPEVAQ--RREQEDLPLYOHQATRVISKASATYGMLSKRAIDTCELP 1048
 QY 261 -----EHLPRFSLDMVTLVEV---PNDGGLGIHVVPFARGRTL-----G 300
 Db 1049 REEGEGEETPNFSGMPRIIVEIFREPNVS--LGISIV-----GGQVIVIRLNGEELG 1101
 QY 301 LVKRLKGGKREHNLRENDCIYRINDGLNRRFEQAQHFRQAMRTPIIF----- 355
 Db 1102 IRIKQVLEDSPAKGNKAKTGDKITLEVSGVDLQASHSEAVEAIKNA-GNPVFTIYQSIS 1160
 QY 356 ---HYVP-----AANKQO-----YQLSQSEKNNYTSR 381
 Db 1161 STPRVTPVANKANKITSNODTQEKKEKROGTAPPMKLPPEYKALDDSDENEEDA 1220
 QY 382 FSPD---SQYD-----NRVNSAGLH----- 400
 Db 1221 FTDOKINGRADLPGEHLITELKDKNGLGLSLAGNKRSMISIVVINGEGRPAADGR 1280
 QY 401 -----TVQADPRLNHP 412
 Db 1281 MHIGDELEINNOILYGRSHQNASALIKTAPSKVAVLPIRNEADAVNMAVTPFPVSSSP 1340
 QY 413 EQLDSHS-----RLHHS-----AHPGSKPPSPAP 435

Db 1341 SSIEDQSTEPTEISEEDGSLVGIKQLPESESEKFLAVSQMKQOKXPTVSSQTEIPLAP 1400
 QY 436 ASAPQNV-----FSTVSSGYNTMKTI--GKRLNQLKKGTEGLCFSTSDVYIGG 484
 Db 1401 ASSYSTADPFTGCGQAPLSDPATCPVPGQEMTIEISGRGGLSLI-----VGG 1454
 QY 485 S---APIYVKNLIPRGALIDGRLKAGDRLLEVNGVDLVKSOBEVVSILRSTKMGTV 540
 Db 1455 KDPPLNAIVIHVEYEGGAARGLMAGDQILEVNGVDLRNSSHEALITLROTPOK--V 1512
 QY 541 SILVFEQDAFPHRELINAPSQMOLPKETKAEDEDIYLTPTGTEFLTEVPYPLNDGSGAG 600
 Db 1513 RLAVYRDE-AHYRDEENLE-----IFPVDLQKKGAG 1543
 QY 601 LGVSVKNSKENHADLGFVKSIIINGGAASKDGLRVNDQLIAVNGESILGKTNDAME 660
 Db 1544 LGLSIYKRRNS-----GVFTSDIYKGGADLDRLTGQDQIILSVNGED-MRNAQSEIYA 1597
 QY 661 TLRSMSTGKNGKMGIOIIVAR-----RISKCNELKSPSPGPPE-----LPLETAL 707
 Db 1598 TLKLC-----AQGLVQLEIGRLRAGSMTSARTTSQNSGSOQSAHSSCHSPFAFVITGL 1651
 QY 708 DD--RERRISHLSYSGIEGLDESPSR---NAALSRIINGS 742
 Db 1652 QNLVGTAKRVSDPSQKN-SGTDMPRIIVEINRELSDALGIS 1690

RESULT 3
 US-09-080-855-12
 ; Sequence 12, Application US/09080855A
 ; Patent No. 6083721
 ; GENERAL INFORMATION:
 ; APPLICANT: Saras, Jan
 ; APPLICANT: Franz, Petra
 ; APPLICANT: Aspenst, Pontus
 ; APPLICANT: Helman, Ulf
 ; APPLICANT: Goner, Leonel Jorge
 ; TITLE OF INVENTION: PARC, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPPL
 ; FILE REFERENCE: L0461/7030
 ; CURRENT APPLICATION NUMBER: US/09/080,855A
 ; EARLIER FILING DATE: 1998-05-18
 ; EARLIER APPLICATION NUMBER: 08/805,583
 ; NUMBER OF SEQ ID NOS: 39
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 2466
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-080-855-12

Query Match 4.7%; Score 332; DB 3; Length 2466;
 Best Local Similarity 20.9%; Pred. No. 1.5e-15;
 Matches 193; Conservative 133; Mismatches 353; Indels 244; Gaps 32;

QY 80 DEODPHGGDGTSSSTGQSPSEIGSELS-----TNVSAFOPYQATSET--E 126
 Db 879 DAODI-----ERASPSRLMQLQASVAGFNMGRALISTGSLASTINKLAVRPLVSQATILK 934
 QY 127 VTPSVLRANMPLH--VRSSDPAL-----IGLSTVSDFSSESESRNPTRMSTTAGE 179
 Db 935 LSCSLSLYQPLQNSSEKKNKDKASWEKPREMSKSYHDLQSASLYHHRN----- 984
 QY 180 IKONTAGSPTCKDRKDENYRSLPR--DTSWMSNOFORDNARSLASHPMVKMKLEKQEO 238
 Db 985 VIVNPEPPPTVALVGVKPSHOMKRSQDAESLAGVYTLNNSKSVASLN-----RSP 1034
 QY 239 DEDGTEEDNSRVEPVGHADTGLHINFSIDM-----VTLVEVPNDG--PL 284
 Db 1035 ERKHESDSSSIDPQQA-----YVLVDYLHKRMSTVSSPEKEITLVNMLKRDARYGL 1085

```

QY 285 GIHVVPESARGRTLLGKLEKGAHEHLEFRENDCIVRINDGDLNRNRFEOAQHMF 344
      1086 GFOIIGEEKMRDLGIFITSSVAPGPGADPHGCLKPGDRLISVNSVSEGVSHHAIETL 1145
QY 345 ROA-----MRPIIMFHV-----VPAANKEQY----- 366
      1146 QNAPEDVTLVISOPEKISKVPSTPVHLTNEMKNYMKSSYMODSAIDSSSKDHMSRGT 1205
QY 367 -----EQLSQSEKNNTYSSRF-----SPDS 386
      1206 LRIHSENFSGPGLREGSLSSQDSRTESASLSQVNGCFASHLDGTQWESQHGSPSP 1265
QY 387 OYIDNRVNSAGLHTVQ---RAPRLNHPPEQID--SHSRLPHSAHPGKPPASAPAPONY 442
      1266 SVISKATEKETFTDSNQSKTKKPGISDVTDYSDRGSDMDDEATYSSQDHQPPKQSSSS 1325
QY 443 FSTYVSSGYNT-----KKIKRLNIQLKKTEGLGFSIT---SRDVTIGSADPIYVKNIL 494
      1336 VNTSKMNFKEFTSSPPKPGDIFEVLELAKNDNSLGI SVTGGVNTSVRRHG---IYVKA VI 1382
QY 495 PGGAIDQGRILKAGDRLIEVNGVDLVGKSOEEVSLRSTKMEGTVSLVFRQEDAFHPR 554
      1383 POGAABESGRIRHKGRVLAIVNGVSLGEGATHKQAVETLRNT---GOVYHLLLEKGGSPITSK 1439
QY 555 E-----LNAEPSQMOIP---KETRAEDEDIVLTPDGTREPLTFEVPPL-NDSGSAG 600
      1440 EHVPTVPOCTLSQNAQOGGPEKVKTTQYKDYSEV-----TEENTFEVKLFKNSSGIG 1493
QY 601 LGVSVKGRSKENHADLGIIFVKSIIINGAASKDGRRLRVNDOLIAVNGESILGKTODAME 660
      1494 FSPSEEDMLIPEQINASTIVRKVLPAGOPAAESGKIDVGVTLKNGASLKGLSQOEYVS 1553
QY 661 TLRRMSTEGNKRGMIOIIVARRISKCNELKSPGPELPPIETALDREERISHSYLS 720
      1554 ALRGAPR-----VFLLICR-----PPGVLPIDRLAL-----LTPLOS 1587
QY 721 GIEGLDESPPRNALSR-----MGESGKYOL-SPTVNMPODDTVIIEEDRLPYL 769
      1588 PAQVLPNKSQSPSCVEQSTSSDENEMSPKSKKQCKSPSRSDYSDSGSGEDLVTA 1647
QY 770 PPHSIDOS-SSSSHDVGFVNADACTMAKAS-----DSADC 806
      1648 PANISNSTWSSALHQTLLNMVSOAQSHHAPRSQEDTICTMYEYPOKIPNKKDEFEDSNPS 1707
QY 807 SLSPDVPVLAFOREGFGROSKS 829
      1708 PLPPMAGQSTYQPSSESASSS 1730

```

```

RESULT 4
PCT-US94-09943-2
Sequence 2, Application PC/TUS9409943
GENERAL INFORMATION:
APPLICANT:
APPLICANT:
APPLICANT:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: TWOMEY, MICHAEL J.
REGISTRATION NUMBER: P-38,349
REFERENCE/DOCKET NUMBER: LC461/7000WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIL
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2466 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
PCT-US94-09943-2

Query Match 4.7%; Score 332; DB 5; Length 2466;
Best Local Similarity 20.9%; Pred. No. 1.5e-15;
Matches 193; Conservative 133; Mismatches 353; Indels 244; Gaps 32;

QY 80 DBODPHHGDDGTSASSTGTQSPETIRGSELG-----INNVSAPFPYQATSTI--E 126
      879 DAODI-----ERASPSRLNLAESVKGFMNGRAISTGSLASSTLKLAVPLSVQAEILRR 934
QY 127 VTPSVLRANMPLH--VRSSDPAL-----IGLSTVSDSNFSSEPSRKNPTRWSTAGF 179
      935 LGCSELSTLQPLONSSKENDKASMEKPREMSKYHDLQSALYHRRN----- 984
QY 180 LKONTAGSPKTCDDRKQDENVRLPR-DTSNMSQFORDNARSLASHPVMKMLEKQDQ 238
      985 VIVNMEPPQTVAEIVGKPSHQMSRSDAESLAGVTKLNSKSVASLN-----RSP 1034
QY 239 DEDGTEDNSRYEPVGHADTGLEHPINFSLDQ-----VKLVEYNDQG-PL 284
      1035 ERKHESSDSSIEDPQA-----YVLVDLHKWSTVSSPERBITLVNLKDKKAYGL 1085
QY 285 GIHVVPESARGRTLLGKLEKGAHEHLEFRENDCIVRINDGDLNRNRFEOAQHMF 344
      1086 GFOIIGEEKMRDLGIFITSSVAPGPGADPHGCLKPGDRLISVNSVSEGVSHHAIETL 1145
QY 345 ROA-----MRPIIMFHV-----VPAANKEQY----- 366
      1146 QNAPEDVTLVISOPEKISKVPSTPVHLTNEMKNYMKSSYMODSAIDSSSKDHMSRGT 1205
QY 367 -----EQLSQSEKNNTYSSRF-----SPDS 386
      1206 LRIHSENFSGPGLREGSLSSQDSRTESASLSQVNGCFASHLDGTQWESQHGSPSP 1265
QY 387 OYIDNRVNSAGLHTVQ---RAPRLNHPPEQID--SHSRLPHSAHPGKPPASAPAPONY 442
      1266 SVISKATEKETFTDSNQSKTKKPGISDVTDYSDRGSDMDDEATYSSQDHQPPKQSSSS 1325
QY 443 FSTYVSSGYNT-----KKIKRLNIQLKKTEGLGFSIT---SRDVTIGSADPIYVKNIL 494
      1326 VNTSKMNFKEFTSSPPKPGDIFEVLELAKNDNSLGI SVTGGVNTSVRRHG---IYVKA VI 1382
QY 495 PGGAIDQGRILKAGDRLIEVNGVDLVGKSOEEVSLRSTKMEGTVSLVFRQEDAFHPR 554
      1383 POGAABESGRIRHKGRVLAIVNGVSLGEGATHKQAVETLRNT---GOVYHLLLEKGGSPITSK 1439
QY 555 E-----LNAEPSQMOIP---KETRAEDEDIVLTPDGTREPLTFEVPPL-NDSGSAG 600
      1440 EHVPTVPOCTLSQNAQOGGPEKVKTTQYKDYSEV-----TEENTFEVKLFKNSSGIG 1493
QY 601 LGVSVKGRSKENHADLGIIFVKSIIINGAASKDGRRLRVNDOLIAVNGESILGKTODAME 660

```

DB 1494 EFSREBNLPEIOINASYVRKKIFAGOPAAESGKIDVDVILKNGASLGLSQOEVS 1553
OY 661 TLRRSMTEGCKRMQIOLIVARISKCNELKSPGSPGPELPIETALDRERISHS 720
DB 1554 ALKGTAE-----VFLLCR-----PPGVLPIDIAL-----LTPIQS 1587
OY 721 GIEGLDESPRNALSHI-----MGESGYOL-SPYVNMPODVTIETDRLEVL 769
DB 1588 PAQVLPNNSDSSQPCVEQSTSSDENEMSKKCKSPSRDYSIDSSGGEIDLTA 1647
OY 770 PPHLSDOS-SSSHDDVGYTADAGTWAKAALS-----DSADC 806
DB 1648 PANISNWTWSALHOTLSNWSQAQSHHEAPKSOEDITCTMYFYPOKIPNKPEFEDSNPS 1707
OY 807 SLSPDVDPVLAFOREGFGROGMS 829
DB 1708 PLPDMAPGOSTOPOSESASSS 1730

RESULT 5
US-08-596-291-3
; Sequence 3, Application US/08596291
; Patent No. 5821075
; GENERAL INFORMATION:
; APPLICANT: GONEZ, LEONEL JORGE
; APPLICANT: SARAS, JAN
; APPLICANT: CLAESON-WEISH, LENA
; APPLICANT: HELDIN, CARL-HENRIK
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596, 291
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/115, 573
; FILING DATE: 01-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: 10461/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; TELEX: 92-1742 EZEKTEL
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2465 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-596-291-3

Query Match 4.7%; Score 327.5; DB 2: Length 2465;
Best Local Similarity 20.9%; Pred. No. 3.2e-15;
Matches 194; Conservative 134; Mismatches 346; Indels 253; Gaps 34;

OY 80 DRQDPHHGGDGTSGASTGTQSPETIFGSELG-----TNNVSAFOPOYATSEI--E 126
DB 879 DAQDI-----ERASFRSLNLAQSVGNFMNGRAISITGSLASTIKLAVPLVSQAEILKR 934
OY 127 YPSPVIRAMPPLH--VRASDPAL-----IGISTVSQSNFSSEEPSKNTTRSTAGF 179
DB 935 LSCSELSTVQPLQNSKSKENDKASWEKPREMSKSYHLSQAISYPRKN-----984
OY 180 LKONTAGSPKTCRDKDENTRSLPR-DTSMNSNOFORNARSLSASHPMVKLEKQEO 238
DB 985 YIVMEPPQTVALELVGKPSHOKSKSDASLAGVYTKLNKSKVASLN-----RSP 1034
OY 239 DEQTEEDNSRVPVGHADTGLEHNPFSIDM-----VKIYEPNNGG-PL 284
DB 1035 ERKHEDSSSIDPQGA-----YLDVLYHKRWSIVSSPERITLVNLTAKAKYGL 1085
OY 285 GIHVYFVSARGR---TLGLLYKRLKLGKKAHEMLFRENQCIYRINDGDLRNRFEQA 340
DB 1086 GFQII-----GGERMETDGLFISVVAQGPADPHGCLKPGDRLISVNSVLEGVSHHA 1140
OY 341 QHMRQA-----MRTPIWPHV-----VPAKNEQY- 366
DB 1141 IETIQNAPEDTLVISQPKKISKVPSTPVHLTWMKMYKKSSYMQDSALDSSSKDHIM 1200
OY 367 -----EQLSQSEKNVYSGRF-----382
DB 1201 SRGTLRHISENSFGPGGLREGSLSSQDSTESASLSQSYNGFASHLGDQTMQESQHG 1260
OY 383 SPDSQYIDNRSVNSAGLHTVQ---RAPRLNHPPOID-SHSRLPHSAHPGKPPSPAPASA 438
DB 1261 SPSPVTSKATKEKFTFDSNGSKTKKPGISDVTYSDGSDMDQEAIVSSQDHQHPKOE 1320
OY 439 POWNFSTVSSGTYN-----KIKGRMLNOLKKGEGGFSIT---SDVYTIIGSADIVY 490
DB 1321 SSSSVNTSKMKMKTFFSSPPKPGDIFVEYELAKNNSIGISVGTGVNYSVHGG---IYV 1377
OY 491 KNILPQGAALODRLKAGDRLLEVNGVDLVGKSOEVVSLRSTMECTVSLVPRODA 550
DB 1378 KDVIPOGALESQDRHKHGDVLAIVNGVSLGATHTKQAVETLNT---GOVHLLLEKGS 1434
OY 551 FHPRE-----LNAEPSQMOIP---KETKADEDIVLPDGTRELFEEVPL-ND 596
DB 1435 PTKSEHVPTPOCTLSQNNQOGGPEKVKTYOVNDYSV-----TEBNFEVALFYN 1488
OY 597 GSAGLGVSVKGNSEKHNHDLGIFVSIINGAASKDGLRPNVDLIANGSLLGKTNQ 656
DB 1489 SGLGFSRREDNLPEQINASYVRKKIFAGOPAAESGKIDVDVILKNGASLGLSQ 1548
OY 657 DAMETLRRSMTEGCKRMQIOLIVARISKCNELKSPGSPGPELPIETALDRERISH 716
DB 1549 EVISALRGTAPE-----VFLLCR-----PPGVLPIDIAL-----LT 1582
OY 717 SLVSGIEGLDESPRNALSHI-----MGESGYOL-SPYVNMPODVTIETDR 765
DB 1583 PLQSPAQVLPNNSDSSQPCVEQSTSSDENEMSKKCKSPSRDYSIDSSGGEIDLTA 1642
OY 766 LPVLPNLSDOS-SSSHDDVGYTADAGTWAKAALS-----D 802
DB 1643 LVTAPANISNWTWSALHOTLSNWSQAQSHHEAPKSOEDITCTMYFYPOKIPNKPEFED 1702
OY 803 SADCSLSPDVDPVLAFOREGFGROGMS 829
DB 1703 SNPSPLPDMAPGOSTOPOSESASSS 1729

RESULT 6
US-09-100-804-3
; Sequence 3, Application US/09100804
; Patent No. 6066472
; GENERAL INFORMATION:
; APPLICANT: GONEZ, LEONEL JORGE
; APPLICANT: SARAS, JAN
; APPLICANT: CLAESON-WEISH, LENA

APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: LC461/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2465 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-100-804-3

Query Match 4.7%; Score 327.5; DB 3; Length 2465;
Best Local Similarity 20.9%; Pred. No. 3.2e-15;
Matches 194; Conservative 134; Mismatches 346; Indels 253; Gaps 34;

QY 80 DEODPHHGDDGTASSTGTQSPETFGSELG-----TNNVSAFOPYQATSET--E 126
DB 879 DAODI-----ERASFRSLNQAESVGRFNMGRATISTGLASTLKLAVRPLSYQAEILKR 934
QY 127 VTPSVLRANPLH--VRSSDPAL-----IGLSTVSDSNFSSEEPKRNPTKSTTAGF 179
DB 935 LSCSELSTLYOPLQNSSKEKNKDKASWEKPREMSKSYHDLQASLTPYPRKN----- 984
QY 180 LKONTGSPKTDCKRDKDEYRSLPR-DTSNMNSQFORDNARSSLASHPRVGVKMLEKOEQ 238
DB 985 VLVNMEPRPQTVAELVGKFSHOSRSDASLAGVTKLNNSKSVASLN-----RSP 1034
QY 239 DEDGTEEDNSRVEPVGHADTGLEHINPFSLDDM-----VKLVEVPNDGG-PL 284
DB 1035 ERKHEHSDSSSTIEDPQA-----YVLVDLHKRMSIVSSPERETITLVNKKDKAYGL 1085
QY 285 GHHVVFSGARGR---TGLGLVLRLEKGGKAEHMLFRENDCIVRINDGLNRPRPEQA 340
DB 1086 GRCIIT-----GGEKMETDLGIFISSVAPGPGPADFHGCLKPGDRILSVNSVSLGVSHHAA 1140
QY 341 OHMFROA-----MRTPLIMFHY-----VPRANKEQY- 366
DB 1141 IETLQNAPEEDVLVISOPEKISKVSPSTVPHVLTNEMKNTWKKSYSMODSALIDSSSKDHW 1200

QY 367 -----EQLSQSEKNNYSSRF----- 382
DB 1201 SRGTLRHISENSFGSGGLREGSLSSQDSRTESASLSQSYVNGFASHLDDQTWQESQH 1260
QY 383 SPDSQYIDNNSVNSAGLHTVQ---RAPRLNHPBQID-SHRLPHSAHPGKPPSPAPASA 438
DB 1261 SPSPSVLSKATKEKETETFDNSQSKTKKPGISDVVDYDSRGSDMDMEAVYSODHQTPKOE 1320
QY 439 PQNVSTVSSGYNT-----KKIGKRLNIQLKGTBELGFSIT---SRDYLIGGSARITY 490
DB 1321 SSSSVNTSNKMNFKTESSSPKPGDIEVELAKNDSLGLSTVGCVNTSVRHG---ITY 1377
QY 491 KNILPRGAALIODGRLKADRLIEVNGVDLVGKSOEEVYSLRSTKMGVTSILVFRQEDA 550
DB 1378 KDVIPOGAASDGRIRHKGDDVLAVNGSLGATHKQAVETLRNT---GVYVHLLLEGQS 1434
QY 551 FHPRE-----LNAEPSOMQIP---KEYAEDEDIVLTPDGTREPLTFEYPL-ND 596
DB 1435 PTKSEHVPVTPQCTLSDONAOGOGPEKVKTKTQVKDYSFV-----TEENTFEVKLFKNS 1488
QY 597 GSAGLGYSVYKNSKEHADLGIYFKSIINGAASKDGRLVNDQLAVNGESLGTKNQ 656
DB 1489 SGLGFSFSREDNLIPEDINASIVKVKLFAGQPAAESGKIDVGDVILKVGASLKLGLSQ 1548
QY 657 DAMETLRRSMSTEGNKGMIQLIYARRISKCNELKSPGSPPELPJETALDDERRISH 716
DB 1549 EYISALRGTAPE-----VFLLCR-----PPGVLPETDNL-----LT 1582
QY 717 SLYSGIEGLDESPSRNALSLRI-----MGESGKYOL-SPTVMPDODVILIEDR 765
DB 1583 PLQSPAGVLPNNSKSDSQPCVEQSTSDENEMSKKCKOCSRSDYSYSSGSGEDD 1642
QY 766 LPVLPRLHSDQS-SSSHDVGFTYADAGTAKAATIS-----D 802
DB 1643 LVYAPANTISWTWSSALHQTLSNWNVSOAQSHHEAPKQEDPTCTMYEYPOKIPKPEFD 1702
QY 803 SADCSLSPDVPYVLAFOREGFGQSMS 829
DB 1703 SNPSPLPRPDMAPOQSYQOSEASASSS 1729

RESULT 7
US-09-290-640-46
; Sequence 46; Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Marcussen, Eric G.
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
; FILE REFERENCE: ISPH-0351
; CURRENT APPLICATION NUMBER: US/09/290,640
; CURRENT FILING DATE: 1999-04-12
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 2485
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-290-640-46

Query Match 4.6%; Score 324.5; DB 4; Length 2485;
Best Local Similarity 20.7%; Pred. No. 5.4e-15;
Matches 193; Conservative 134; Mismatches 361; Indels 245; Gaps 32;

QY 80 DEODPHHGDDGTASSTGTQSPETFGSELG-----TNNVSAFOPYQATSET--E 126
DB 879 DAODI-----ERASFRSLNQAESVGRFNMGRATISTGLASTLKLAVRPLSYQAEILKR 934
QY 127 VTPSVLRANPLH--VRSSDPAL-----IGLSTVSDSNFSSEEPKRNPTKSTTAGF 179
DB 935 LSCSELSTLYOPLQNSSKEKNKDKASWEKPREMSKSYHDLQASLTPYPRKN----- 984

```

QY 180 LKONTAGSPKTCRRKDEYNSLR-DTSMNSQFQDMARSSLSASHPWGKMLEKQ 238
Db 985 VYINMEPPQVIAELVGRPSHOMKSDSESLAGVTKLNKSVASLN-----RSP 1034
QY 239 DEDGTEENSRVEPYGHA-----DTGLEHIFPESLDM-----VKLV 275
Db 1035 ERKHESSSSSIEDPGQAVYLGMTMHSSGSSQVPLKENDVLRKMSIVSSPERETLV 1094
QY 276 EYPNDGG-PIGIHVVPFSAAGRTGLVLRLEKGGKAKHEMLFRENDCIYRINDGDRN 334
Db 1095 NLKMDAXYGLGFOIIGGKGRIDLGIPTLSSVAPGGPDLDCCLPGRLLISVNSYLEG 1154
QY 335 RRFQAOHMFROA-----MRPTIMEHV-----VPAA 361
Db 1155 VSHHALEILQNAPEQVTVLVIQPEKLSKVPSTVHLTNMKMYMKSSMODSADISS 1214
QY 362 NKREY-----EOLSOSEKNNTYSSRF-----382
Db 1215 SKDHHSKGLNRHISENSFGPSGGLREGSLSSQDSRTESLSQSYNGFASHLGDQTM 1274
QY 383 -----SPDSQYIDNRSVNSAGLHTVQ---RAPLNMHPQID-SHRLPHSAHPGKRP 432
Db 1275 QESQHSFPSPVYSKATKEFTTDSNOSKTKKPGISDVTDYSDRGSDMDEATYSSODH 1334
QY 433 SAPASAPQNVSTYVSSGYNT-----KKIGKRLNIOLKKGTEGTSIT---SRDVTIGG 484
Db 1335 QTPKQESSSVNTSKMNFETSSSPKPGDIEVELAKNDNSLGI5VYGCVN5VRHGG 1394
QY 485 SAPYVKNILPRGAIDGRKAGDRLEIVNGVDLVKSGOEVSLLSTKMEGTVALY 544
Db 1395 ---YVNAVILPOGAESDGRHKGDRVLAIVNGVSLGATHKQAVETLNT---QGVVHL 1448
QY 545 FRODAFHRE-----LNAEPSOMQIP---KETKAEDEDIYLPDSTREPLTFEEV 591
Db 1449 LEKQSPSTKEHVPYTPQCTLSDQNAOGGPEKVKTTQVADYSFV-----TEENFEV 1502
QY 592 PL-NDGSAGLVSVKGNRSKENHADLGIFVKSIIINGAASKDGRLEVNOLDIYANGESI 650
Db 1503 KLEKNSGGLGFSPSREDNLIPEQINASTIVRYKKLPQCPAPAESGKIDVGIILKVGASL 1562
QY 651 LGKTNODAMETLRMSSTEGKRGKIQILTVARISKCNELKSPSPGPELPTETRLDNR 710
Db 1563 KGIISOEVIALRGTAPE-----VFLLCR-----PPPGVLEIDRL---1600
QY 711 ERRIHSILXSGIEGIDSPSRNALISRI-----MGESKTYOL-SPTVNMPODDTV 759
Db 1601 ---LTPQSPAQVLPNSSKSDSQPCVEGSTSSDENKSDSKCKSPSRHDSYSDSS 1656
QY 760 IIEDRLPVLPRPHLSDOS-SSSSHDDYGVYTLADAGTMAKAALIS-----801
Db 1657 GSGEDDILVTAPANISNTSSALHQTLSNMVSOAQSHHEAPKSOEDTICTMFFYYPKIPN 1716
QY 802 -----DSADCSLSPDVPYLAQREGFGROGMS 829
Db 1717 KPFEEDSNPSPLPPDMAPOGYOQSESASSS 1749

```

```

RESULT 8
US-08-410-804-1
; Sequence 1, Application US/08410804
; Patent No. 5632994
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive, Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States

```

```

ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/410,804
FILING DATE: 27-MAR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/259,514
FILING DATE: 14-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1389
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-410-804-1

```

```

Query Match 3.9%; Score 275.5; DB 1; Length 610;
Best Local Similarity 24.8%; Pred No. 2.7e-12;
Matches 132; Conservative 77; Mismatches 193; Indels 131; Gaps 19;
QY 360 AANKEQYEQLOSSEK-----NTYSSRFSPDSQYIDNRSVNSAGLHTVQAPRLNMHP 412
Db 12 ATEKEFTTDSNOSKTKKPGISDVTDYSDRGSD---MDEATYSSQDH-----QTP 59
QY 413 EOLDSIRLPHSAHPGKRPAPASAPQNVSTYVSSGYNTKKIGKRLNIOLKKGTEGIG 472
Db 60 KQ-ESSSVNTSKMNFETSSSPKPGDIP-----EVELAKNDNSLG 101
QY 473 ESIT-----SRDVTIGGSAPIYVKNILPRGAIDGRKAGDRLEIVNGVDLVKGSQ 524
Db 102 ISVTYLFKGGVNTSVRAGS---IYVAVILPOGAESDGRHKGDRVLAIVNGVSLGATH 158
QY 525 EEVVALSTMEGTVALVFRQDAFHRE-----LNAEPSOMQIP---KETKA 571
Db 159 KOAVETLRNT---QGVVHLLLEKQSPSTKEHVPYTPQCTLSDQNAOGGPEKVKTTQV 215
QY 572 EDEDIYLPDGTREFLTFEYPL-NDGSAGLVSVKGNRSKENHADLGIFVKSIIINGAA 630
Db 216 KDYSFV-----TEENFVKYKLFKNSGGLGFSPSREDNLIPEQINASTIVRYKKLPQCPA 269
QY 631 SKDGRLEVNOLDIYANGESLKGKTNODAMETLRMSSTEGKRGKMIQILTVARISKCNEL 690
Db 270 AEGKIDVGIILKVGASLKGISOEVIALRGTAPE-----VFLLCR-----314
QY 691 KSPGSPGPELPTETRLDNRERRISHLSYSGIEGIDSPSRNALISRI-----MG 740
Db 315 ---PEPGVLEIDRL---LTPQSPAQVLPNSSKSDSQPCVCPQSSDENKMS 363
QY 741 ESGKYOL-SPTVNMPODDTVIIEDRLPVLPRPHLSDOS-SSSSHDDYGVYTLADAGTMAKA 798
Db 364 DKSKQCKSPSRHDSYSDSSGSGEDDILVTAPANISNTSSALHQTLSNMVSOAQSHHEA 423
QY 799 AIS-----DSADCSLSPDVPYLAQREGFGROGMS 829
Db 424 PKSOEDTICTMFFYYPKIPKPFEDSNPSPLPPDMAPOGYOQSESASSS 476

```

```

RESULT 9
US-08-259-514-1
; Sequence 1, Application US/08259514

```

Patent No. 5747245
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Sato, Takaaki
TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cathryn Campbell
STREET: 4370 La Jolla Village Drive, Ste 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/259,514
FILING DATE: 14-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9954
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-259-514-1

Query Match 3.9%; Score 275.5; DB 1; Length 610;
Best Local Similarity 24.8%; Pred. No. 2.7e-12;
Matches 132; Conservative 77; Mismatches 193; Indels 131; Gaps 19;

QY 360 AANKEQYEOLOSSEKN-----NYSSRFPDSQYIDNRSVNSAGLHTVQAPRLNHP 412
DB 12 ATEKEFTDSDNGSKTKRPGISDVTDYSDRGDSD--MDEATYSSODH-----QTP 59
QY 413 EQDISHRLPHSAHPGKPPSAPASAPONVFSTVSSGYTKIKGRRLNIQLKKGTEGLG 472
DB 60 KO-ESSSVNTSKRMKMFPTSSPPKPGDIF-----EVELAKNDNSLG 101
QY 473 FSIT-----SRDYTIGSAPYYVKNILPRGAIDGRKAGDRLEIENGVDLVKRSQ 524
DB 102 ISVTVLFDKGCVNTSVRHG---IYKAVIPQGAESDGRHKGDRVLAVNGVSLGATH 158
QY 525 EEVYSLRSTKMEGTIVSLVFRQEDAFHPR-----LNAEPSOMQIP---KETKA 571
DB 159 KQAVETLRNT---GQVYHLLLEKGSPTSKEHVPVTPQCLSDQNAOGSPREYKKTQY 215
QY 572 EDDDIYVLPDGTRELFTEFVPL-NDSGSAGLVSVKGNRKENHADLGIYVKSIIINGCA 630
DB 216 KDYSFY-----TEENTFEYKLFKNSSGIGFSTRSDNLIPEOINASYVKKLFPQPA 269
QY 631 SKDRLRVNDQLAVNGESILGKTNDAMETLRRSMSTEGNKKGMQLIYARRISKNEL 690
DB 270 AESGKIDVGIVILKVGASIKGLSQGEVISALNGTAP-----VFLILCR----- 314
QY 691 KSGSPGPGPELPTETALDDRERRISLSLIGLDESPRNALSTI-----MG 740
DB 315 ---PPGVLPETIDAL-----LTPLOSAPQVLPNSSSKDSSQPCVCEOSTSDENEMS 363
QY 741 ESGKYL-SPYVMQDDDTYIEDDRLPVLPRLSDQS-SSSSHDVGFYTAGAGTWAK 798
DB 364 DSKKCKCKSPSRDSTSDSGSGEDLVYAPANISNTWSSALHQTLSNMVSOQASHHEA 423

QY 799 AIS-----DSADCSLPDVPVLAFOREGGRQSMS 829
DB 424 PKSQEDTICMFEYYPQIKRPFEDSNPSPLPDMAPGSGYQGESASSSS 476

RESULT 10
US-08-858-311-1
Sequence 1, Application US/08858311
Patent No. 5876939
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Sato, Takaaki
TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cathryn Campbell
STREET: 4370 La Jolla Village Drive, Ste 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,311
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/410,804
FILING DATE: 27-MAR-1995
APPLICATION NUMBER: US 08/259,514
FILING DATE: 14-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1389
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 610 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-858-311-1

Query Match 3.9%; Score 275.5; DB 2; Length 610;
Best Local Similarity 24.8%; Pred. No. 2.7e-12;
Matches 132; Conservative 77; Mismatches 193; Indels 131; Gaps 19;

QY 360 AANKEQYEOLOSSEKN-----NYSSRFPDSQYIDNRSVNSAGLHTVQAPRLNHP 412
DB 12 ATEKEFTDSDNGSKTKRPGISDVTDYSDRGDSD--MDEATYSSODH-----QTP 59
QY 413 EQDISHRLPHSAHPGKPPSAPASAPONVFSTVSSGYTKIKGRRLNIQLKKGTEGLG 472
DB 60 KO-ESSSVNTSKRMKMFPTSSPPKPGDIF-----EVELAKNDNSLG 101
QY 473 FSIT-----SRDYTIGSAPYYVKNILPRGAIDGRKAGDRLEIENGVDLVKRSQ 524
DB 102 ISVTVLFDKGCVNTSVRHG---IYKAVIPQGAESDGRHKGDRVLAVNGVSLGATH 158
QY 525 EEVYSLRSTKMEGTIVSLVFRQEDAFHPR-----LNAEPSOMQIP---KETKA 571
DB 159 KQAVETLRNT---GQVYHLLLEKGSPTSKEHVPVTPQCLSDQNAOGSPREYKKTQY 215
QY 572 EDDDIYVLPDGTRELFTEFVPL-NDSGSAGLVSVKGNRKENHADLGIYVKSIIINGCA 630

```

Db 216 KDYSFV-----TEENTFEVLFKNSSGIGFSFREDLILPEQINASIYRKKLFPGQPA 269
QY 631 SKDRLRYNDOLIVANGSILGKTNOAMETLRBSMSDEGKRGMIOLIVARLSKCNEL 690
Db 270 AESGIDGVDYLKANGSLGSOQEVTSALRGAP-----VFLDLR----- 314
QY 691 KSPSPPEPELPIETALDRERRISLSYSGIEGIDESPSSRNALSR-----MG 740
Db 315 -----PPPEVLIEDLAL-----LTPAQSPAQVLPNSSKSSQSPSCVEQSTSDENEMS 363
QY 741 ESGYOL-SPVYNNPQDDTVIIEDRLPVLPPHSDOS-SSSHDDVCFVADAGTAKA 798
Db 364 DKSKKQCKSPSRSDYSDSGSGEDDLVTAPANISNSTWSSALHQTLSNMVSOAQSHHEA 423
QY 799 AIS-----DSADCSLSPVDVPLAFOREGFCRGSMS 829
Db 424 PKSQEDTCTMYFYFQKIPNKEFEEDSNPSPLPDMAGSYQPOSESASSS 476

```

```

RESULT 11
US-08-545-860D-48
; Sequence 48, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the A11-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESS: No. 6040140ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1612 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-545-860D-48

Query Match 3.7%; Score 260; DB 3; Length 1612;
Best Local Similarity 20.5%; Pred. No. 1.7e-10;
Matches 297; Conservative 190; Mismatches 450; Indels 510; Gaps 77;

QY 49 HRLHGGGILDDLDLIGDV-----ADKDRLYAVFDEODPH----- 86
Db 307 HSDKGAKEIT-LDDDECPLOIFREWPSDKGILVFLQKRPPDHPKTKKHLBEGTKPKG 365
QY 87 -GGDGTSSAST-----GTQS---PEIRG-----SELGT-----NNVSAF 116
Db 366 KERADSYGSGTLPEKPLVLELSPGSDSRDKRYRLQLSTVEVTEKLDNSIGLF 425
QY 117 -----QPYA-----TSELYTPSVIRAMPPLHVRSSPALI--GLSTVSDSN-FSSE 164
Db 426 GPGIOPHHCIDLINMDGVVTVPRSDAETVEGGRISTMTMQSGMKQFGASHVFKFVD 485
QY 165 PSRKNP-TRWGTAGFLKONTAGSP-----KTCDRKD-ENYRSLPRTSWMNSNOFQDN 217
Db 486 PSQDHALLAKRSDGGLMKVGRPHRGIVQETTFDGLGDIHSGTALP--TSKSTRILDSBR 543
QY 218 ARSSLASH-----PMVG-----KMLEKQEDDGT-----EDNSRVEP 252
Db 544 VSSASTAERGVMKPMIRIVEGOQPDYRQESRTQDASGPELLIPASIEPSESSESLSAI 603
QY 253 VGHADTGLEHT---PNFSLDMVKLYEVPNDGGPLGIHVYFSSARGGTLLILYKRLKG 309
Db 604 INTNNSSTVHFKLSPYVLYMACRYV-LSNOYRP--DISP-TEETHVIVAVNMV--- 655
QY 310 GKAHEHNLFRENDCIV-----RINDGDLRRRFEQAOHMRQAMRPIIMFHVYPAANK 363
Db 656 ---SMMEGYIQOKNTAGALAFMMANASSELIN-FIKQDSLSRITLDADQVLAHLVQMAFK 712
QY 364 EOYEQLSQSEKNNYSSRFPSPDSOYIDNRSVNSAG-----LHTVO-----RAVRLNH 410
Db 713 YLVHCL-QSELNMYPA-----FLDDPEENSLOPRKIDYVLAHTLTAMSLIRRCRYN- 763
QY 411 PPEQIDSHRSLPHSAHPGKRPASAPASAPQVNESTVS---SGYTKKIGKRLNTOL--- 464
Db 764 AALTIQLPFSQLPHFTN-----MWLEFNLVDPDSGLCSHWGALIIROQLDHI 810
QY 465 -----KKGTGEGSITSRDV-----TIGSAPATYVNI-----LPGAIADGR 504
Db 811 EAMAKQGLLELAADCHLSRIYQATLLMDKYPADIDININSTCKKMSLOLQALQYVH 870
QY 505 LKAGRLIEVNGVDLVGKSQEEVSVLSRSTKMEGVSLVYRQEDAFHPRLENAFSPSQM 564
Db 871 CAPDEPFIPT-TDLI-----ENVVVAVENTADE-----LARS DG---REVOLR- 908
QY 565 IPKTKAEDEDIVL-----TPDGRREFLT----- 588
Db 909 -----EDPDQLPFLLPEDGYSCDVVRNIPNGIQEFLDPLCQGFRLILPHTRSPGTW 961
QY 589 ---FE-----VPLNDSGSAGLG--VSYKGNRSKENHADL 617
Db 962 TIYFEGADYESHLENTLEAQLAKKEPILITVILKKNQNGLSIYAAKG-----AGQK 1017

```



```

Db 909 -----EDPDLQFLPEBDGYSDVVRNIPNGIQEFLDPLQGRGFCGLIPHRSPGTW 961
QY 589 ---FE-----VPLNDSCSAGLG-VSKGRNSKENHADL 617
Db 962 TTYFEGADYESHLEKRENTIEAOLRKEPELITVTLKKONGKLSIVAAGK---AGODKL 1017
QY 618 GTFVKSIIINGAASKDGLRLVNDOLIVANGESILKTNODAMETLRSMST---EGNKRK 674
Db 1018 GIYKSVYKGAADVDRKLAAGDOLLSVDRSLVGLSOERAAELMTSTSSVITLLEVAAGK 1077
QY 675 -----MIOLIVARISKCNELKSP-----GSPGPEL----- 701
Db 1078 AIYHGLATLLNOSPMPMQRISDRSGSKPRKSEGFELYNNTQNSPSPQOLPMAEYSE 1137
QY 702 PIETALDDR--ERRISHSLYSGIEGLDESPSRNAALSRIENGSGKQLSPTVNM----- 753
Db 1138 PKRLPDDDLKMKNRADHSSSPVANOPSPSGKSAVA--SGTTAKTIVSGNLCTEBQT 1195
QY 754 --PODDTVI-----EDRLPVL-----PHLSOSSSSSHD 784
Db 1196 PPPREAVPIPTQYTYTREYTFPPASKSODRM--APPQONMPNYYEERKPHMHTSDHSS--- 1250
QY 785 VGFYTAAGTWAKAALISDSADCSLSPDVDPVLAFOREGFGQSMSEKRTQFSDASQIDP 844
Db 1251 -----TALQRTVRSQEBLREDAKQOL-ERRHIEA 1278
QY 845 VKTRKSKMDIGIADETKL--NTVDDQKAGSPSDVGPSTGLKKS-----SLESIQ 894
Db 1279 AMDRKSDS--DMWINGSSSIDSTSSQEBLNHSSKSVTPASTLTKSGPGWKTPTAIPATP 1337
QY 895 TAVAEVTLNGDIPFRPRRIIRGRGCSFRAAIDKSDKRAVDDDEGMEETLEEDPEE 954
Db 1338 VAVSQ-PIRTDLPPLPPPPV-----HYAGFDMSMDLLPPLPPS 1377
QY 955 SRSRGREVSRTASDQPSHSLEROMNGNOKGDKTDR--KKDTGKKEKKDKKDKKDMA 1012
Db 1378 ANQILPSAOVAA-----AERKREHQRWTEKEKALEPERER----- 1416
QY 1013 KKGMLKGLGDMFRGKHRRKDKIERTGKTIQ-----ESFTSEERIRIMQOERRIOA 1065
Db 1417 -----KRRQER--KLGMKQSLNPAFSPILAAQ--MKPEKPSITLQ 1455
QY 1066 KTRB--FRROARENDYAEIODFHRTGCDDELMTGVSSESGMALNAR----- 1114
Db 1456 RPOETVIRELQPOQ-----RTERRD-LQYITVSKBELSSGDSLSPPMKRDAKE 1506
QY 1115 --OSPEGHMDALYAOVKKPRNSKPSPYDNSRSTSNDRIOQLROEF-----GQAKQ 1166
Db 1507 KLEKQOQHIIYDMLSKETIOE-LQSKP-----DRSAEES--DRLKILMEMQOKLDESKO 1559
QY 1167 -DEVED 1172
Db 1560 KDEDDEE 1566

```

RESULT 13

```

US-09-157-420-1
Sequence 1, Application US/09157420
Patent No. 6180760
GENERAL INFORMATION:
APPLICANT: TAKAI, Yoshimi
APPLICANT: NAKANISHI, Hiroyuki
APPLICANT: MANAI, Kenji
APPLICANT: WADA, Manabu
TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-AFADIN"
FILE REFERENCE: 98-10442/LC(PAC)/653
CURRENT APPLICATION NUMBER: US/09/157,420
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1829

```

```

; TYPE: PRT
; ORGANISM: rat
US-09-157-420-1

Query Match
Best Local Similarity 3.78; Score 260; DB 4; Length 1829;
Matches 179; Conservative 131; Mismatches 329; Indels 272; Gaps 35;

QY 530 ILRSTKMEGYVSLVVRQEDAFHPRLENAEPSONQIFRETKAEDDEDIVLPDQTRREFLTF 589
Db 974 LVPHTRSPGTWTFYF--EGADYESHLEKRENTIELOLRKEPE-----VIT----- 1016
QY 590 EYPLNDSCSAGLG-VSKGRNSKENHADLGFVKSIIINGAASKDGLRLVNDOLIVANGE 648
Db 1017 -VTLKKONGKLSIVAAGK---AGODKLGIYKSVKGAADVDRKLAAGDOLLSVDR 1071
QY 649 SLTGKTNODAMETLRSMST--EGNKRK-----MIOLIVARISKCNEL 690
Db 1072 SLVGLSOERAAELMTSTSSVITLLEVAAGKQAIYHGLATLLNOSPMPMQRISDRSGSKPR 1131
QY 691 KSP-----GSPGPEL-----PIETALDDR--ERRISHSLYSGIEGLDESPS 730
Db 1132 KSEGEFLYNNSAONGSPSPQOLPMAEYSEPKRLPDDDLKMKNRADHSSSPVANOPSPSG 1191
QY 731 RNALSLINGESGKQLSPTVNM-----PODDTVI-----EDR 765
Db 1192 GKSPYT--SGTTAKTIVSGNLCTEBQTPPREAVPIPTQYTYTREYTFPPASKSODRM 1249
QY 766 LPVL-----PHLSOSSSSSHDVGFTADAGTWAKAALISDSADCSLSPDVPLA 817
Db 1250 APVQONMPNYYEERKPHMHTSDHAS-----IA 1275
QY 818 FOREGFGQSMSEKRTQFSDASQIDPVTKTRKS--MDIGIADETKNTVDDQKAGSPSR 876
Db 1276 IQRTVRSQEBLREDAKQOL-ERRHIEA-----SLESIQ 894
QY 877 DVGPSGLKKS-----SLESIQTAVA-EYTLNGDIPFRPRRIIRGRGCSFRAAI 929
Db 1335 SVTASLTITLTSGGKRWKTPAAVLPTPVAASQPIRTDLPPLPPPP-----PAHY 1382
QY 930 DKSTDKPAVDDDEGMEETLEEDTESSRSRGRESVSTASDQPSHSLEROMNGQEGDRTD 989
Db 1383 TSQFDISMD-----LPLPPPAQAAPQSAOVAERKREHQRWTEKERA--- 1430
QY 990 RKDKTGKREKDRKQKAKKGMKGLGDMFRGKHRRKDKIERTGKTIQESFTS 1049
Db 1431 -RLEERERKREDEKRLGQKRTQSLNPASFSLTAQAKPEKPSITLQ-----P 1478
QY 1050 EERIRBMQEOERTQARTREFRROARENDYAEIODFHRTGCDDELMTG--GVSYEGSM 1108
Db 1479 QETVIRELQPOO--QPTTERRDLO-----YTTIS-----KEELSSGDSLSPPMKR 1523
QY 1109 ALNARPOSPEGHMDALYAOVKKPRNSKPSPYDNSRSTSNDRIOQLROEF-----Q 1162
Db 1524 DAREKLEKQOQHIIYDMLSKETIHELQN-----KGDRTAEEDDRKLTMLMOPQKRLQ 1576
QY 1163 QAKODEVDDRRTYSFQDPWPNARPATOSGRHSYSEVOMOROROEHESSQAOQOYS 1222
Db 1577 ESKQDEDEDEED-----DDVDTMLIIMORLAEARRALOQDEERRQ 1618
QY 1223 SLPRQSRKASVSQDSMEQNYSPGEGFOSAKENPRYSYQSGRNQYIGGFGNARMLE 1282
Db 1619 QOLEMKRR--VEDRVROEF-----GRHOREERKRD 1650
QY 1283 TOELLROBO--RRKEQOMKOPSPSGSPNDYSYKQVDP-----SYAPKGP 1327
Db 1651 AEERKROEEGYRLEARRRO-----HEEAARLLEPEEGGLSRPPLPDYEPSPS 1703
QY 1328 FRODVPPSPSQ 1338
Db 1704 SPSAPPPPPQ 1714

```


CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,632
FILING DATE: 19-MAR-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/041,016
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Corless, Peter F.
REGISTRATION NUMBER: 33,860
REFERENCE/DOCKET NUMBER: 48147/1699-CIP
TELEPHONE: 617-523-6440
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1061 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-32

Query Match
Best Local Similarity 3.2%; Score 222.5; DB 3; Length 1061;
Matches 196; Conservative 134; Mismatches 362; Indels 335; Gaps 41;
Db 215 RDNAKSSLSASHPMVGKMLEK-QEDEDCTEEDNSRVEPV-GHADTGLEHIFNPSL----- 268
116 RGAHDDRNKSRPVYITCRPGPDRETGKPGDRLSLVDGIRLGTTHAEMSLKQC 175
269 -DMVATVE-----VPNDGPLGIHVVPFSARGRTLG-----LIVKRL 306
176 GQEAATLITDYVSAANDSVATASGPL--LVEVAKTPGASLGLVLTISVCCNKQIVIDKI 232
307 EKGGAHEHNTFRENDCIYRINDGDLRNRREQA-----QMFROAMR 349
233 KASAIADRCALHVGDHILSDTSMEXCTLAETQFLGNTTDOVKLEILPHQTRLAK 292
350 TPIWFHVVPANKQEQYQLSOSSEKNYSSRPSDQYIDNRSVNSAGLHVQARPLN 409
293 GP-----DHVKIQRSDRLPM-----DPMASQCSVHT-----NHH 323
410 HPPQIDSHSLPHSAHPSGKPPSAP-----ASAPQNVSTVS-----SGYNTKK 455
324 HNPHPD-HCRVPALGFPKALFPNSPPAWSSSPYSMSAYSLSLNMGTLPRSLYSP 382
456 IGRRLNTQLKGTGEGFTTSRDVTIG----- 484
383 RGTMMRRRLKKKPKSSLSLASTVGLAGVYHTEVTLADPYTGFIQLOGSVFAT 442
485 ---SAPLYKNIIPGCAIYDRLKAGDRLBNGVDLVGKSQEVVSLRSTMECTVS 541
443 ETLSPLSLYLEADSPAERCGLQIGRVAMINIGIPTEDSTFEANQLRDSITSKYT 502
542 L-----LVPRQDAFH-----PRELNAE-----PSOMQ-----IPKETA----- 571
503 LEIEPVAESVTPSSGTFHVKLPPKKHVELGITISSPSRKPDPPLVYISDIKKGSVAHRT 562
572 -----ED-----EDIY-----LTPDGTRE-----FLTFEV 591
563 GTLELGDKLAIIDNIRLDSMEDAVOILDOCEDLVKLTKRDEDSNDRQESSGAIITYV 622
592 PLNDGSAGLGVSVKNSKENHADLGIYKSIINGASAKDRLRVNDOLIAVNGESIL 651
623 ELKRYGSP-LGITISGT---EPPD-PIIISLTGKGLAERTGALHIGRIALINSSSLK 677
652 KKTODAMETLRSMSTEGNKKRGMTOIIVARRISKNELEKSPSPGPELPETALDRE 711
678 GKPLSEDIHLQWAGER-----VTLKIKKOTDAQPASSP---KKLPFP----- 717
712 RLISHSLYSGIEGLDESP-SRNAALSR-----MGESKYYQLSPTYNM 753

Db 718 ---SHSDLGGEEDPSPFIORPGKLSDPVSTVPSVSDAVSDWDSGIDARCGSGCTTFQ 774
Qy 754 PODDTVILEEDRLVLPPLHS--DQSSSSHDVDGVTADACTWAKAISADCSLSPD 811
Db 775 TSGYNFTYWRSPKRRASLSPVKKPRSQTYPDVGLSND---WDRSTASGFAGASDSAD 831
Qy 812 VDPVLAFOREGF----- 828
Db 832 AE-----QEENFWSQALEDELTCCGSGILRELEATIMSGTMSLNHAPTPARQLGQAS 886
Qy 829 SEKTKQFSDASQLDVYTKRKSXM--DLGIADETKLTNYD--OKAGSPRDVGPISGLK 885
Db 887 FQERSN-----SRPHYSQTRNSNTLPDVGKRSYTLRKMGEIETMSPTPVELHKTLY 941
Qy 886 KSSLSLSIQTAVAEYTLN-----GDIPIFRPRPRIIRGRCNESFRAAIDKS 932
Db 942 KDSGMEDEFGFVADGULEKGYVYANIRPAGPDGLGKLPYDRLLQ---VHVTRTDFDCC 998
Qy 933 YDKPAVDDEGEMTLEEDTESSRSGRRESYSTASD-----QPSHSLERQNMQ 982
Db 999 LVVPLIASGKNKLDIVISRNPDIASOKSIEQALPDSWMSBQNSAFPOQPSH-----GGNL 1052
Qy 983 EKGDKTD 989
Db 1053 ETREPTN 1059

Search completed: July 24, 2002, 14:30:46
Job time: 7488 sec


```

1183 TCGAGAACAAAGCAGACACCTCCTGCCCGTGGAAACACAGCTCTGAAGGC 1232
416 pSerHisSerArgLeuProHisSerAlaHisProSerGlyLysProPro 433
    ||| : : : : : ||| : : : : :
1233 AGTACAT.....CTAACAGAGGAGCAGCCCTGAAGAGAAAGAGCAG 1276
433 erAlaPro..... 435
    : : : : :
1277 CCTCCCGCAGCAAGCAGAGTCCCGGGTTCACAGACTAGAGAGAAA 1326
436 AlSerAlaProGlnAsnValPheSerThrValSerSerGlyTyrAs 452
    ||| : : : : : ||| : : : : :
1327 CCTTCCTCTCC.....TCACCTCTCTCTCTCATGGGGTTTGG 1364
452 nThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGlyThrg 469
    : : : : : : : : : : : : : : : : : : : : : :
1365 CAGCAAGAAAATAATGCAAAAAATAATTAAGATTGACCTAAAGAAAGCCCTG 1414
469 IuclLysLeuGlyPheSerIleThrSerArgAspValThrIleGlyLys 485
    ||| : : : : : ||| : : : : :
1415 AGGGACTTGGTTTCACTGTGTGTAACAGAGATTCTCTATACATGGTCT 1464
486 AlAProIleTyrValLysAsnIleLeuProArgLysAlaAlaIleGln 502
    : : : : : : : : : : : : : : : : : : : : : :
1465 GGTCCCATTTTGTAAAAACATCTTACAAAAGAGCAGCAGTAAAGCA 1514
502 p.GlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyValAsp 518
    ||| : : : : : ||| : : : : :
1515 TGACC...CTACAAATCAGGGGACAGAAATTTTGGAGTAAATGGCAGAGAT 1561
519 LeuValGlyLysSerGlnGlnGluValValSerLeuLeuArgSerThrLys 535
    : : : : : : : : : : : : : : : : : : : : : :
1562 GTTACAGGAAGAACCAAGAAAGATCGTGCCATGTTAAGAGACACAA 1611
535 smetGlnGlyThrValSerLeuLeuValPheArgGlnGluAspAlaPhe 552
    ||| : : : : : ||| : : : : :
1612 GCAGGGAGACACAGCTGATTTGCTCATTTGCCCGCAGAGAGGAGCTTCC 1661
552 IsProArgLysLeuAsnAlaGluProSerGlnMetGlnIleProLysGlu 568
    ||| : : : : : ||| : : : : :
1662 TTGCCGAGAGCTGAAGAGAACTGATGTATGCTCTCTCCCTGAG 1711
569 ThrLysAlaGlnAspGluAspIleValLeuThrProAspGlyThrArgL 585
    : : : : :
1712 TCAAGC.....GA 1719
585 uPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLeu 602
    ||| : : : : : ||| : : : : :
1720 GCAGCTTCACCTTGGAGATCCCTCGAATGATTTCTGGCTCCGCTGGCTGG 1769
602 LysAlaSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuGly 618
    ||| : : : : : ||| : : : : :
1770 GGGTTAGCCTGAAGAGGAATTAAGTCTCGAAGAACTGGAAGTGAAGTGGG 1819
619 IlePheValLysSerIleIleAsnGlyLysAlaAlaSerLysAspGlyArg 635
    ||| : : : : : ||| : : : : :
1820 ATTTTATCAATTCATCATCATCGAGGTGCTCTTCAAGAGTGGCCG 1869
635 GluArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeuG 652
    ||| : : : : : ||| : : : : :
1870 TCTCCCAATATATGACCACTGATTTGCTTAAAGGGGAAACTCTTCTGG 1919
652 LysThrAsnGlnAspAlaMetGlnThrLeuArgArgSerMetSerThr 668
    ||| : : : : : ||| : : : : :
1920 GAAAGTCAACACAGAAAGCATGAAGAAACACTCAGGCGATTCATTCATG 1969
669 GlnGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg.... 683
    ||| : : : : : ||| : : : : :
1970 GAGGAAACATCCGGGGCATGATCCAGTTGTGATCTGAGAGAGACCTGA 2019
684 .....IleSerLysCysAsnGluLeuLysSerProGly 695
    : : : : : : : : : : : : : : : : : : : : : :
2020 GAGACCACTGGAGAGCTTTCAGAAATGGAGACATTTCAGACCAAGCT 2069

```

```

695 erProProGlyProGlnLeuProIleGlnThrAlaLeuAspArgGlu 711
    ||| : : : : : ||| : : : : :
2070 TTGAGAACTGTCAGAA.....GCTGTAGACACTCCAG 2104
712 ArgArgIleSerHisSerLeuTyrSerGlyIleGlnLysLeuAspGlu 728
    ||| : : : : : ||| : : : : :
2105 CGAAATGATAGTACTATATTTATAT..... 2128
728 rProSerArgAsnAlaAlaLeuSerArgIleMetGlyLysSerGlyLys 745
    ||| : : : : : ||| : : : : :
2129 .....CCGTTTGGCAGCT 2141
745 YrgLysLeuSerProThrValAsnMetProGlnAsp.....AspThr 758
    ||| : : : : : ||| : : : : :
2142 ATACT.....CCACAAGATAAAGAAAGACCTA 2170
759 ValIle.....IleGluAspAspArgLeuProValLeuPr 770
    : : : : : : : : : : : : : : : : : : : : : :
2171 TTGCTTCCAGTGTATGGTGGCTGAGAAATGAACTACACCCGCTCCGCC 2220
770 oProHis.....LeuSerAspGlnSerSerSers 781
    ||| : : : : : ||| : : : : :
2221 ACCACATCTCTCTCTCGAATGGGGCTTTGAAGATTTCAGTCACGCTCTG 2270
781 erHisAspAspValGlyPheValThrAlaAspAlaGlyThrTrpAlaLys 797
    : : : : : : : : : : : : : : : : : : : : : :
2271 GAGTGATACACACAGATATTT..... 2293
798 AlaAlaIleSerAspSerAlaAspCysSerLeuSerProAspValAspPr 814
    ||| : : : : : ||| : : : : :
2294 .....CCAGAT..... 2299
814 oValLeuAlaPheGlnArgGlnGlyPheGlyArgGlnSerMetSerGlu 831
    ||| : : : : : ||| : : : : :
2300 .....CAACATCAACTTC..... 2314
831 ysArgThrLysGlnPheSerAspAlaSerGlnLeuAspPheValLysThr 847
    : : : : : : : : : : : : : : : : : : : : : :
2315 .....AGAACGTGACACACCACTCAGCAGCCGGAATGATTAACCTT 2356
848 ArgLysSerLysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAs 864
    : : : : : : : : : : : : : : : : : : : : : :
2357 AAAGCTTCCAAGACATGAGACTT...GTGCCAATGCAAGCAAGTTCA 2403
864 nThrValAspAspGlnLysAlaGlySerProSerArgAspValGlyPro 881
    : : : : : : : : : : : : : : : : : : : : : :
2404 GTCCCTGGCGGATGCGAGATCAGACTCTCCAGGCAAGATTTTGTCCAA 2453
881 erLeuGlyLeuLysLysSerSerSerLeuGlnSerLeuGlnThrAlaVal 897
    : : : : : : : : : : : : : : : : : : : : : :
2454 CTTTGGGCTGTAAGAAAGTCCAGTTGCTTGGAGACCTACAGACGGCTGTG 2503
898 AlaGluValThrLeuAsnGlyAspIleProPheHisArgProArgProAr 914
    ||| : : : : : ||| : : : : :
2504 GCTGAACTCAGAGAAAT...GATCTGCCCTTCCACAGGCCCAACACCA 2550
914 GileIleArgGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspL 931
    : : : : : : : : : : : : : : : : : : : : : :
2551 CATGTGCGGGGCGGAGGCTCCACAGAGAGCTTCCGAGACCCATCCACA 2600
2601 AATGTCATGATGGCCAGAA...GAGGACAGCGGATGTGCTG.... 2641
948 LeuGlnGluAspThrGlnGluSerSerArgSerGlyArgGlnSerValSe 964
    : : : : : : : : : : : : : : : : : : : : : :
2642 .....TCTGATTAACAGCTCTGCTCAGGCCACACAGCTTTG.. 2677
964 rThrAlaSerAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyA 981
    : : : : : : : : : : : : : : : : : : : : : :
2678 .....AATTGTGAGTCTGCCCCCTCAGGGA 2702

```

```

981 snGlnGluYsGlyAspLysThrAspArgLysLysAspLysThrGlyLys 997
    |||
    |||
    |||
2703 ACCCCGAGGTTA...GACAAATGTGAAATAAACCAAAAACATCAAAAAA 2749
998 GlnYsLysLys 1001
    |||
    |||
    |||
2750 ACAAAACAGAG 2761

seq_name: gb_est1:AL529631

seq_documentation_block:
LOCUS      AL529631                                1078 bp    mRNA
DEFINITION AL529631 lT1_NFL001_NBC4 Homo sapiens cDNA clone CS0DD005YD18.3
prime: mRNA sequence.
ACCESSION  AL529631
VERSION    AL529631.1 GI:12793124
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 1078)
            Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
    source
        1..1078
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DD005YD18"
            /clone_lib="lT1_NFL001_NBC4"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed
            by life technologies. Contact : Feng Liang Medical
            Technologies, a division of Invitrogen 9800 Medical Center
            Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
            8371 Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"

BASE COUNT      236 a      283 c      222 g      331 t      6 others
ORIGIN
alignment_scores:
    Quality: 1600.00      Length: 359
    Ratio: 4.776          Gaps: 4
    Percent Similarity: 93.315      Percent Identity: 90.251

alignment_block:
US-09-757-781-2 x AL529631/rev ..
Align seg 1/1 to reverse of: AL529631 from: 1 to: 1078

696 ProProGlyProGluLeuProIleGluThrAlaLeuAspAspArgGluAr 712
    |||||
    |||||
    |||||
1063 CCCCCCTGGACCTGGCTGCCATTGAKACACGCTTGAGTGAAGAAG 1014
712 GARGIleSerHisSerLeuTYrSerGlyIleGluGlyLeuAspGluSerP 729
    |||||
    |||||
    |||||
1013 AAGAATTTCCATTCCCTACAGTGGGATGAGGGCTTGATGAATCGC 964
729 rGserArgAsnAlaAlaLeuSerArgIleMetGlyGluSerGlyLysTYr 745
    |||||
    |||||
    |||||
963 CCAGCAGAAATGCTGCCTCAGTAGATGAATG.....GGTAATAC 923

```

```

746 GlnLeuSerProThrValaMetProGlnAspAspThrValIleIleG1 762
    |||||
    |||||
    |||||
922 CRGCTGCCCTACAGTAATATGCCCCAAGATGACACGTGATTATAGA 873
762 uAspAspArgLeuProValLeuProProHisLeuSerAspGlnSerSers 779
    |||||
    |||||
    |||||
872 AGATGACAGGTGCCAGTGCTTCCTCCACATCTCTGACCGCTCTCTT 823
779 eRserSerHisAspAspValGlyPheValThrAlaAspAlaGlyThrTrp 795
    |||||
    |||||
    |||||
822 CCGGCTCCCATGATGATGGGGTTGTGACCGCAGATGCTGCTACTGG 773
796 AlAlaYsAlaAlaIleSerAspSerAlaAspCysSerLeuSerProAspVa 812
    |||||
    |||||
    |||||
772 GCCAAGGTGCATATCAGTATTCACGCCGCTCTTTGAGCCAGATGT 723
812 lAspProValIleuAlaPheGlnArgGluGlyPheGlyArgGlnSerMetS 829
    |||||
    |||||
    |||||
722 TGATCCAGTTCCTTGGCTTTTCAACGAGAAAGATTGGACGTGAGATATGT 673
829 eRgLnLysArgThrLysGlnPheSerAspAlaSerGlnLeuAspPheVal 845
    |||||
    |||||
    |||||
672 CAGAAACGACACAAAGCAATTTTCAAGTCCAGCATATGGATTTGGT 623
846 LysThrArgLysSerLysSerMetAspLeuGlyIleAlaAspGluThrLy 862
    |||||
    |||||
    |||||
622 AAAACACAAATAATCAAAAAGCATGATTTA...GTGCTGACGAGACTAA 576
862 sLeuAsnThrValAspAspGlnLysAlaGlySerProSerArgAspValG 879
    |||||
    |||||
    |||||
575 ACTCAATACATGATGATGACCAAGAAAGGATCTCCMGACAGAGATGTGG 526
879 lYrProSerLeuGlyLeuLysLysSerSerSerLeuGlnSerLeuGlnThr 895
    |||||
    |||||
    |||||
525 GTCCCTCCCTGGCTCTGAGAAAGTCAAGCTCGTTGAGATCTGCAGACC 476
896 AlAlaValAlaGluValThrLeuAsnGlyAspIleProPheHisArgProAr 912
    |||||
    |||||
    |||||
475 GCAATTCGCCAGTGTGATTTGAATGGGATATTCCTTCCATCGTCCAGC 426
912 gProArgIleIleArgGlyArgGlyCysAsnGlnSerPheArgAlaAlaI 929
    |||||
    |||||
    |||||
425 GCCCGGATATCATCAGAGCAGGGATGCAATGACRCTTCAGGCGCTGCCA 376
929 lAspLysSerTYrAspLysProAlaValAspAspAspGluLysMet 945
    |||||
    |||||
    |||||
375 TCGACAAATCTTATGATTAACCCGCGTAGATGATGATGATGAGGACATG 326
946 GlnThrLeuGlnGluLysAspThrGlnLysSerArgSerArgGlyLysE 962
    |||||
    |||||
    |||||
325 GAGACCTTGGAGAGAGACACAGAAAGTTCAAGATCAGGAGAGAGATGC 276
962 rValSerThrAlaSerAspGlnProSerHisSerLeuGlnArgGlnMetA 979
    |||||
    |||||
    |||||
275 TGTATCCACAGCCAGTATGATCAGCTTCCACCTCTGTGAGAGACAAAGA 226
979 snGlyAsnGlnGluLysGlyAspLysThrAspArgLysLysAspLysThr 995
    |||||
    |||||
    |||||
225 ATGGAAACCAAGAGAAAGGTGATTAAGACTGATAGAAAAAAGATTAAC 176
996 GlyLysGlyLysLysAspArgAspArgAspLysGlyLysAspLysMetLysAl 1012
    |||||
    |||||
    |||||
175 GGAAGAAAGAAAGAAAGAAAGATGAGATTAAGAGAAAGATTAAGAAAG 126
1012 AlLysLysGlyMetLeuLysGlyLeuGlyLysMetPheArgPheGlyLysH 1029
    |||||
    |||||
    |||||
125 CAAGAAGGAAATGCTGAAGGCTTGAGGACATTTCAAGCTTGCCAAA. 77
1029 lAsArgLysAspAspLysIleGluLysThrGlyLysLysIleGlnGlu 1045
    |||||
    |||||
    |||||
76 ..CTGAAGCCCGAGAGAGATGAGACAACAAAGCGATTCAAA..... 38

```

1046 SerpetherSergluGlucLuarGlie 1054
||| |||:|||||
37ACATGTTTGAACAGCACATA 17

seq_name: gb_est1:AL529864

seq_documentation_block:
LOCUS AL529864 1074 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529864 LTI_NFL001_NBC4 Homo sapiens cDNA clone CSDD005YH02 3
prime, mRNA sequence.
ACCESSION AL529864
VERSION AL529864.1 GI:12793357
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1074)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD005YH02"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 229 a 289 c 221 g 324 t 11 others
ORIGIN

alignment_scores:
Quality: 1496.00 Length: 347
Ratio: 4.603 Gaps: 4
Percent Similarity: 93.660 Percent Identity: 88.473

alignment_block:
US-09-757-781-2 x AL529864/rev ..

Align seg 1/1 to reverse of: AL529864 from: 1 to: 1074

696 ProProGlyProGluLeuProIleGluThrAlaLeuAspAspArgGluAr 712
||||| ||| |||:|||||
1042 CCCCCCTGACCTGGTCCCATG...ACACGCTGATGATGAGACG 997
712 gATGATSerHisSerLeuTySerGlyIleGluGlyLeuAspGluSerP 729
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
996 .ACAAATTCCTTCTTACAGTGGAGTTGAGGGCTTGATGATGATCGC 947
729 roSerArgAsnAlaAlaLeuSerArgIleMetGlyGluSerGlyLysTy 745
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
946 CCAGCAGAAATGCTGCCCTCAGTAGGATTAATG.....GGTAATAC 906
746 GluLeuSerProThrValaAsnMetProGluAspAspThrValaIleIle 762
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
905 CMGCTGCTCCCTACAGTGAATATGCCCCAAGATGACACTGTCTATTAGA 856

762 uaAspAspArgLeuProValaLeuProProHisLeuSerAspGlnSerSers 779
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
855 AGATGACAGAGTTCACAGTCTCTCCACATCTCTGACACAGCTCTT 806
779 erSerSerHisAspAspValGlyPheValThrAlaAspAlaGlyThrTrp 795
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
805 CCVAGCTCCATGATGATGGTGGTGTGTGACGACAGATGCTGTACTTGG 756
796 AlAlaValAlaAlaIleSerAspSerAlaAspCysSerLeuSerProAspVa 812
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
755 GCCAAGGCTCAATCACTGATTCAGCCGACTGCTTGTGATCCAGATGT 706
812 laAspProValaLeuAlaPheGlnArgGluGlyPheGlyArgGlnSerMetS 829
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
705 TGATCCAGTCTCTGCTTTTCACAGAGAAGATTGGACGTCAGAGTTGGT 656
829 erGluLysArgThrLysGlnPheSerAspAlaSerGlnLeuAspPheVal 845
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
655 CAGAAANACGACAAACCAATTTTCAGATGCCACTCAATTGGATTGCTT 606
846 LysThrArgLysSerLysSerMetAspLeuGlyIleAlaAspGluThrLys 862
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
605 AAAACACGAAATCAAAAACCATGAT...TTWCTTCTGACGAGACTAA 559
558 ACCTCAATACAGTGTGATGACCAAGAAAGAGGCTTCCCGCAGAGATGTGG 509
879 LysProSerLeuGlyLeuLysSerSerSerLeuGlnLysLeuGlnThr 895
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
508 GTCTCTCTCCGTGGTCTGAAGAAAGCTGCTGGTGGAGAGCTGCAAGCC 459
896 AlValaIaGluValaThrLeuAsnGlyAspIleProPheHisArgProAr 912
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
458 GCAGTTCGACAGAGTGTGAATGGGATATTCCTTCCATGCTCCACG 409
912 gProArgIleIleArgGlyArgGlyCysAsnGluSerPheArgAlaAlaI 929
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
408 GCCCGGATATATCAGAGCGAAGGATGCAATGAGCGCTTCAGGCTGCCA 359
929 leAspLysSerTyraAspLysProAlaValaAspAspAspGluGlyMet 945
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
358 TCGCAATCTTWTGATTAACCCCGGTAGATGATGATGAAGAGGATG 309
946 GluThrLeuGluGluAspThrGluGluSerSerArgSerGlyArgGluSe 962
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
308 GAGACCTTGGAGAGACACAGAAAGAAATTCAAGATCAGGAGAGAGTC 259
962 rValSerThrAlaSerAspGlnProSerHisSerLeuGluArgGlnMetA 979
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
258 TGTATCCACAGCCAGTATGATGAGCTTCCACTCTCTGAGAGCAAAATGA 209
979 snGlyAsnGlnGluLysGlyAspLysThrAspArgLysLysAspLysThr 995
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
208 ATGGAAACCAAGAGAAAGGTGATTAAGTGTAAACCAAAAGATTAAGACT 159
996 GlyLysGluLysLysLysAspArgAspLysGluLysAspLysMetLysAl 1012
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
158 GGAAAAAGAAAGAAAGAAAGATAGAGATAGAGAGAAAGATTAATAAGC 109
1012 AlLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgPheGlyLysH 1029
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
108 CAAGAAAGGAATGCTGAAGGCTTGGAGACATGTTGACGCTTGCCAAA. 60
1029 tsArgLysAspAspLysIleGluLysThrGlyLysIleLys 1042
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
59 ..CTGAAGCCCGAGAGAAGATGACAACAAGCGATTCAAA 21

seq_name: gb_est2:BG745599

seq_documentation_block:
LOCUS BG745599 852 bp mRNA linear EST 15-MAY-2001

```

DEFINITION      602723832f1 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:4850526 5',
                  mRNA sequence.
ACCESSION       BG745599
VERSION         BG745599.1 GI:14056252
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 852)
AUTHORS         NIH-MGC http://mgc.ncl.nih.gov/
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                  Email: cgabs-r@mail.nih.gov
                  Tissue Procurement: Dr. Mark Watson
                  CDNA Library Preparation: Ling Hong/Rubin Laboratory
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LNL at:
                  http://image.jnl.gov
                  Plate: LICM1691 row: 0 column: 07
                  High quality sequence stop: 829.
                  Location/Qualifiers
                    1..852
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4850526"
                     /clone_lib="NIH_MGC_113"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCAAGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH-MGC library."
BASE COUNT      237 a 191 c 222 g 202 t
ORIGIN
alignment_scores:
  Quality: 1317.50      Length: 300
  Ratio: 4.739          Gaps: 6
  Percent Similarity: 92.667 Percent Identity: 90.667

alignment_block:
US-09-757-781-2 x BG745599 ..
Align seg 1/1 to: BG745599 from: 1 to: 852

```

```

732 snAlaAlaLeuSerArgIleMetGlyIuSerGlyIysTyrGlnLeuSer 748
|||||
251 ATGTCGCCCTCAGTAGGATTAATG.....GGTAAATACAGCTGTCC 291
|||||
749 ProThrValAsnMetProGlnAspAspThrValIleIleGluAspAsp 765
|||||
292 CCTACACTGCAATATCCCCCAAGATCACACTGCAATTAAGAAGACACG 341
|||||
765 GLeuProValLeuProProHisLeuSerAspGlnSerSerSerSer 782
|||||
342 GTTGCAGTAGCTTCTCCACATCTCTGACGACGACCTCTTCCAGCTCC 391
|||||
792 IsAspAspValGlyPheValThrAlaAspAlaGlyThrTyrAlaLysAla 798
|||||
392 ATGATGATGTGGGGTGTGTGACGGCAGATGCTGTACTTGGGCCAAGGCT 441
|||||
799 AlaIleSerAspSerAlaAspCysSerLeuSerProAspValAspPro 815
|||||
442 GCAATCAGTGATTCAGCCGACGCTCTTGAAGTCCAGATGTTGATCCAGT 491
|||||
815 IleuAlaPheGlnArgGlyGlyPheGlyArgGlnSerMetSerGlyLys 832
|||||
492 TCTTCTCTTTCACAGCAGAAAGCATTTGGACGTACAGTATGTCAGAAAAC 541
|||||
832 rGthrLysGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArg 848
|||||
542 GCACAACAACAATTTTCAGTGCAGTCAATGATGATTTGTTAAACACAGA 591
|||||
849 LysSerLysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsn 865
|||||
592 AAATGAAAAGCATGATGATTAAGT..... 615
|||||
865 rValAspAspGlnLysAlaGlySerProSerArgAspValGlyProSer 882
|||||
616 .....AGTTCTCCACAGACAGATGCGGCTCTTCC 646
|||||
882 euGlyLeuLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAla 898
|||||
647 TGGGCTCTAAGAAAGTCAAGCTCATGTGAGAGTCTCACACCCGCACTGCC 696
|||||
899 GluValThrLeuAsnGlyAspIlePro.PheHisArgProArgProArg 914
|||||
697 CGAGGTGACTTTGAATGGGATATTCTTTTCCATGCTCCAGCGCGCGG 746
|||||
915 IleIleArgGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLys 931
|||||
747 ATCTTCAGAGCGCAGGGGATGCAATGACAGCTTCACAGCTGCCATCGACAA 796
|||||
931 sSerTyrAspLysPro.AlavalAspAspAspAspGluGlyMet 945
|||||
797 ATCTTATGATTAACCGCGGT...AATGATGATGATGAAGGCAATG 837
|||||

seq_name: gb_est2:BM462581
seq_documentation_block:
LOCUS      BM462581 1063 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6426369 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5518290
5', mRNA sequence.
ACCESSION  BM462581
VERSION    BM462581.1 GI:18511621
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1063)
AUTHORS    NIH-MGC http://mgc.ncl.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1AM12178 row: n column: 19
 High quality sequence stop: 365.

FEATURES
 source
 Location/Qualifiers

1. 1063
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5318290"
 /clone_lib="NIH_MGC_71"
 /tissue_type="Telomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: PCMV-SPORE; Site: 1; Note:
 Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.1 kb."
 BASE COUNT 307 a 247 c 259 g 250 t
 ORIGIN

alignment_scores:
 Quality: 1307.00 Length: 363
 Ratio: 4.059 Gaps: 16
 Percent Similarity: 88.705 Percent Identity: 84.298

alignment_block:
 US-09-757-781-2 x BM462581 ..

Align seg 1/1 to: BM462581 from: 1 to: 1063

575 AspIleValIeuThrProAspGlyThrArgGluPheLeuThrPheGluVal 591
 |||||
 1 GATATGTTCTTACACCTGATGACACAGGATTTCTGACATTGAGT 50
 591 lProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValLysGlyA 608
 |||||
 51 CCCACTTAATGATTCAGATCTGCAGGCCCTGGTGTCTCAGTCAAGGTA 100
 608 snArgSerLysGluAsnHisAlaAspLeuGlyIlePheValLysSerIle 624
 |||||
 101 ACCGGTCAAAAGGAAACACGCGATTTGGGAAATCTTTGTCAAGTCCATT 150
 625 IleAsnGlyValAlaIleSerLysAspGlyArgLeuArgValAsnAspGly 641
 |||||
 151 ATTAATGAGAGACGACATCTAAAGATGGAAGGCTCGGGTGAATGATCA 200
 641 nIleuIleAlaValAsnGlyLysSerLeuLeuGlyLysThrAsnGlnAspA 658
 |||||
 201 ACTGATAGCAGTAATAGAGAAATCCCTGTGGGCAAGACAAACCAATATG 250
 658 lAmcLthrIleuArgArgSerMetSerThrGluGlyAsnLysArgGly 674
 |||||
 251 CCATGGAAACCCTAAGAGGCTATATCTCTAAGGCAATTAACGAGGA 300
 675 MetIleGlnLeuIleValAlaArgArgIleSerLysCysAsnGluLeuGly 691
 |||||
 301 ATGATCCAGCTTATGTGCAAGAGAAATTAAGCAAGTCAATGAGCGAA 350
 691 sSerProGlySerProProGlyProGluLeuProIleGluThrAlaLeuA 708
 |||||
 351 GTCACCTGGAGGCCCTGGACCTGAGCTGAGCCCATTTGAACAGCGTTGG 400
 708 spAspArgGluArgArgIleSerHisSerLeuThrSerGlyIleGluGly 724
 |||||
 401 ATCATAGAGAACGAAATTTCCCATTTCCCTCTACAGTGGAGTGAAGGG 450
 725 lLeuAspGluSerProSerArgAsnAlaAlaLeuSerArgIleMetGlyGly 741
 |||||
 451 CTGATGAATACGCCCAAGAAATGCGCCCTCAGTAGATTAATG.... 495
 741 uSerGlyLysTyrGlnLeuSerProThrValAsnMetProGlnAspAsp 758

|||||
 496GGTAAATACAGCTGTCCTTACAGTGAATATGCCCAAGATGACA 541
 758 hrValIleIleGluAspArgLeuProValIleProProHisLeuSer 774
 |||||
 542 CTGCTATTATAGAGATGACAGGTTGCCAGTGTCTCTCCACATCTCT 591
 775 AspGlnSerSerSerSerHisAspAspValGlyPheValThrAlaAs 791
 |||||
 592 GACACAGTCCCTTCCACCTCCATGATGATGTGGGTTTGTGACGCGAGA 641
 791 pAlaGlyThrTrpAlaLysAlaAlaIleSerAspSerAlaAspCysSerL 808
 |||||
 642 TGGTGTGACTTGGGCCCAAGCT.GCAATCAGTATTCAGCCGACTGCT 690
 808 euserProAspValAspProValLeuVal.PheGlnArgGluGlyPheG 824
 |||||
 691 TTGAGTCCGATGGTGAATCAGTCTTGTCTTTCACAGAAAGATTTGG 740
 824 YArgGlnSerMetSerGlu.LysArgThrLysGln.PheSerAspAla.S 840
 |||||
 741 ACGTACAGATATGTCAAGAAAAAGCCAAAGCAATTTTCAGATGCCCA 790
 840 erGlnLeuAspPheVal.LysThrArgLysSerLysSerMetLsp.LeuG 856
 |||||
 791 GCCAATTTGATTTCTTTAAACCCCAAAATCCAAAGCCGTGAAATTAG 840
 856 lIleIleAlaAspGluThrLysLeuAsnThrValAsp..AspGlnLysAla 871
 |||||
 841 GGATAGCTTGAAGACTAACTCATATCTGGATGGCCCCCAAAACCA 890
 872 GlySerProSerArgAspVal...GlyProSerLeuGlyLeuLysLys 887
 |||||
 891 GGCTCTCCCAACCAAAATGTGGGGCTTCCTTGGGTCTGAATAAAA 940
 887 rSerSerLeu.....GluSerLeu.GlnThr.....AlaValAla 898
 |||||
 941 GTCAAACCTCTTGTAGAAAGTGTGGCAAAACCCCAATTTGGCCCAAGG 990
 899 GluValThrLeuAsnGlyAspIleProPheHis...ArgProArg...Pr 913
 |||||
 991 GCACCTTTAAAGGGGAAAAATTCCTTTTCAATCGGCCCCAGGCCCC 1040
 913 oArgIleIle...ArgGlyArg 919
 |||||
 1041 GGGGATTAATTCACCAAGGCCCGG 1062
 seq_name: gb_est2:BI917860
 seq_documentation_block:
 LOCUS BI917860 767 bp mRNA linear EST 16-OCT-2001
 DEFINITION 603183990F1 NIH_MGC_121 Homo sapiens CDNA clone IMAGE:5247718 5',
 mRNA sequence.
 ACCESSION BI917860
 VERSION BI917860.1 GI:16199788
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 767)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail@nln.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: ILAM11625 row: d column: 23
 High quality sequence stop: 759.
 Location/Qualifiers

FEATURES

source

1. 767
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5247718"
 /clone_11b="NIH_MGC_121"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC library."
 BASE COUNT 236 a 179 c 190 g 162 t
 ORIGIN

alignment_scores:
 Quality: 1296.00 Length: 255
 Ratio: 5.082 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 99.608

alignment_block:

US-09-757-781-2 x BI917860 ..
 Align seg 1/1 to: BI917860 from: 1 to: 767

```

373 GULYSASnAsnTYrTyrSerArpPheSerProAspSerGlnTyrll 389
|||||
3 GAGAGAAACAATTACTATTCAAGCCGTTTACCCCTGACAGCCAGTATAT 52
|||||
389 eAspAsnArGSerValAsnSerAlaGlyLeuHisThrValGlnArGalap 406
|||||
53 TCACAAACAGAGAGTGTGAACAGTGCAGGGCTTCACACGGTGCAGAGAGC 102
|||||
406 rArGLeuAsnHisProProGluGlnIleAspSerHisSerArgLeuPro 422
|||||
103 CCCGACTGACACACCCGCTGAGAGATGACTCTCACCACAGACACTACCT 152
|||||
423 HisSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaPr 439
|||||
153 CATAGGCGACACCCCTCGGGAAACACCATCCGCTCCAGCCTCGGCACC 202
|||||
439 oGlnAsnValPheSerThrThrValSerSerGlyTyrAsnThrLysIysI 456
|||||
203 TCAGAGATGATTTAGTACGACTGTAAAGCAGTGTATTAACCAAAAAA 252
|||||
456 lAcGlyLysArgLeuAsnIleGlnLeuLysGlyThrGluGlyLeuGly 472
|||||
253 TAGGGAAGAGCGCTTAATATCCAGCTTAAGAAAGATACAGAAAGTTTGG 302
|||||
473 PheSerIleThrSerArgAspValThrIleGlyGlySerAlaProIleTy 489
|||||
303 TTCAGCATCATCTCCAGAGATGTAAACAATAGGTGGTCCAGCTCCAAATCT 352
|||||
489 rValLysAsnIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuL 506
|||||
353 TGTGAAAAAATTCCTCCCGGGGGGGCCCATTTGAGATGGCCGACTTA 402
|||||
506 yAlaGlyAspArgLeuIleGluValAsnGlyValAspLeuValGlyLys 522
|||||
403 AGCGAGAGAGACAGACTTATAGAGTAAATGAGTAGATTAGGGGCAAA 452
|||||
523 SerGlnGluGluValValSerLeuLeuArgSerThrLysMetGluGlyTh 539
|||||
453 TCCCAAGAGAGAGAGTGTGCTGCTTGAGAGAGACCAAGATGAGAGAAC 502
|||||

```

```

539 rValSerLeuLeuValPheArgGlnGluAspAlaPheHisProArgGluL 556
|||||
503 TGTGAGCCCTTCGTCTTCGCCAGAGAGAGCCCTTCACCAAGAGGAAAC 552
|||||
556 eAsnAlaGluProSerGlnMetGlnIleProLysGluThrLysAlaGlu 572
|||||
553 TGAATGCGAGACCAAGCAGATGCGAGATTCAAAAAGAAACCAACAGAA 602
|||||
573 AspGluAspIleValLeuThrProAspGlyThrArgGluPheLeuThrPh 589
|||||
603 GATGAGGATATCGTCTTCACACCTGATGCGACAGAGGAATTCGACATT 652
|||||
589 eGluValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValL 606
|||||
653 TGAATCCCACTTATGATTCAGATTCGACAGCCTGTGTGTCAGTGTCA 702
|||||
606 yAGLYASnArGSerLysGluAsnHisAlaAspLeuGlyIlePheValLys 622
|||||
703 AAGGTAAACCGGTCAAAAGAAACACGACGAGATCGGGAATCTTGTCAAG 752
|||||
623 SerIleIleAsnGly 627
|||||
753 TCATATTACTCTGCG 767
|||||

```

seq_name: gb_esc2:BM466811

seq_documentation_block:

LOCUS BM466811 1150 bp mRNA linear EST 05-FEB-2002
 DEFINITION AGENCOURT_6431318 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5503072
 5', mRNA sequence.

ACCESSION BM466811
 VERSION BM466811.1 GI:18515853
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 1150)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: ILAM12142 row: d column: 17
 High quality sequence start: 9
 High quality sequence stop: 652.

FEATURES

source

Location/Qualifiers
 1. 1150
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5503072"
 /clone_11b="NIH_MGC_67"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."
 BASE COUNT 294 a 310 c 309 g 235 t 2 others
 ORIGIN

alignment_scores:
 Quality: 1241.00 Length: 246
 Ratio: 5.107 Gaps: 3
 Percent Similarity: 98.780 Percent Identity: 98.374

alignment_block:

US-09-757-781-2 x BM466811 ..

Align seg 1/1 to: BM466811 from: 1 to: 1150

```

1112 AIAATGProGlnSerProArgGluGlyHisMetMetAspAlaLeuTyrAl 1128
|||||
39 GGTAGACCCAGAGCCAGAGAGAGGCGATGATGATGATGATGATGATGATG 88
1128 acGlnValLysProArgAsnSerLysProSerProValAspSerAsnA 1145
|||||
89 TCAAGTCAAGAACCCGCGAATTCCTCAACCTCTCAGCTAGACAGTAACA 138
1145 rGserThrProSerAsnHisAspArgLleGlnArgLeuArgGlnGluPhe 1161
|||||
139 GATCAACTCTCTAGCAATCATGATCGGATGACAGCTCTGAGCAGAAATT 188
1162 GlnGlnAlaLysGlnAspGluAspValGluAspArgArgArgThrTyrSe 1178
|||||
189 CAGCAAGCAAGCAAGATGATGATGATGATGATGATGATGATGATGATG 238
1178 rPheGlnGlnProTrpProAsnAlaArgProAlaThrGlnSerGlyArgH 1195
|||||
239 TTTTGAGCAACCTCGCGAGACGACGCGCGGCGAGCGAGCGGCGGAC 288
1195 lAsserValSerValGluValGlnMetGlnArgGlnArgGlnGlnGluArg 1211
|||||
289 ACTCGGTGTCGTGGAGGTGAGATGACAGCGGCGAGCGGCGAGGAGAGCC 338
1212 GluSerSerGlnGlnAlaGlnArgGlnTyrSerSerLeuProArgGlnSe 1228
|||||
339 GAGACTCTCCAGCAGGCGCGAGCGGCGAGTACAGCTCTCTCTCGGCGAAG 388
1228 rArgLysAsnAlaSerSerValSerGlnAspSerTrpGlnGlnAsnTyrS 1245
|||||
389 CAGGAAAAATGCCAGCTCGCTCTCCAGAGACTCTTGCGAGCGAAGACTACT 438
1245 ePProGlyGluGlyPheGlnSerAlaLysGluAsnProArgTyrSerSer 1261
|||||
439 CCCCTGGGGAGGCTCCAGAGTGCCTCAAGAGAACCCAGAGTCTCTCAGC 488
1262 TyrGlnGlySerArgAsnGlyTyrLeuGlyGlyHisGlyPheAsnAlaAr 1278
|||||
489 TACCAAGGCTCCAGGAGCGGTACCTGAGGAGGAGGAGGCTTCAACCCAG 538
1278 gValMetLeuGluThrGlnGlnLeuLeuArgGlnGlnArgArgLysG 1295
|||||
539 GGTCACTGCGAAACTCAGAGAGCTCTTCCGACGAGAACAGAGGCGGAGAG 588
1295 lGlnGlnMetLysLysGlnProProSerGlnGlyProSerAsnTyrAsp 1311
|||||
589 ACGACAGATGAGAAAGACAGCTCTCTCCGAGGGGCCAGCAACTATGAC 638
1312 SerTyrLysLysValGlnAspProSerTyrAlaPro.ProLysGlyProP 1328
|||||
639 TCGTATTAAGAAAGTCCAGAGCCCACTTACGCCCTCCAGGGGGCCCT 688
1328 heArgGlnAspVal.ProProSerProSerGlnValAlaArgLeuAsnAr 1344
|||||
689 TCCGGCAAGATGTGCCCCCTCTCTCTCAAGTTGCGAGGCTGAGACAG 738
1344 gLeuGlnThrProGlu.LysGlyArgProPhe 1354
|||||
739 ACTTCAGACTCTGAAAAAAGGAGGCCCTTT 770
1354 ACTTCAGACTCTGAAAAAAGGAGGCCCTTT 770
seq_name: gb_est2:BE792557
seq_documentation_block:
LOCUS BE792557 725 bp mRNA linear EST 20-SEP-2000
DEFINITION 60158534.f1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3939370 5',
mRNA sequence.
ACCESSION BE792557

```

```

VERSION BE792557.1 GI:10213755
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at: image.lnl.gov
Plate: L10M788 row: b column: 11
High quality sequence stop: 725.
FEATURES
source
1..725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3939370"
/clone_id="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

253 AAGACACAGAGAAAGTTCAGATCAAGGAGAGAGTCTGTATCCACAGCC 302
967 SeraspGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnG 983
303 AGGATTCAGCCCTCTCCATCTGTGAGAGACAAATGATGAGAAACCAAGA 352
983 ULysGlyAspPlyThrAspArgLysLysAspPlyThrGlyLysGluLys 1000
353 GAAAGGCGATGAAGCTGATGAGAAAGTAAAGTAAAGTAAAGTAAAGTAA 402
1000 yslYsAspArgAspPlyGluLysAspPlyMetLysAlaLysLysGlyMet 1016
403 AGAAAGATGAGATGAAGCAAGCAAGATGAAGTAAAGTAAAGTAAAGTAA 452
1017 LeuLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAsp 1033
453 CTGAAAGGCGCTTGGGAGACATGTTCAAGTGTGGCAACATCGAAAGATGA 502
1033 pLysIleGlyLysThrGlyLysIleLysIleGlnGluSerPheThrSerG 1050
503 CAAGATTGAGAAACGGGTAAATAAATAAATACAGAAATCTTACATCAG 552
1050 LngLngLngArgLleArgMetLysGlnGlnGlnGlnArgIleGlnAlaLys 1066
553 AAGAGAGAGAGATGACATGACAGACAGACAGACAGATTCAGAGCCAAA 602
1067 ThrArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGlu 1083
603 ACTCGAATTTAGGAGACAGACAGCTCGAGAGCTGACTATGTGAAAT 652
1083 eGlnAspPheHisArgThrPheGlyCysAspAspGluLeuMetLysGlyG 1100
653 TCAGATTTTCATCGACATTTGG.TGTGATGATGATGATGATGATGATG 701
1100 LysAlaSerSerTyrGluGlySer 1107
702 GAGTCTCTTATGAGAGTTCC 724

seq_name: gb_est2:BG745723

seq_documentation_block:
LOCUS BG745723 1040 bp mRNA linear EST 15-MAY-2001
DEFINITION 60273832T1 NIH_MGC_113 Homo sapiens CDNA clone IMAGE:4850526 3',
mRNA sequence.
ACCESSION BG745723
VERSION BG745723.1 GI:14056376
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1040)
NIH_MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabrs@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
Plate: LNCM1681 row: 0 column: 07
High quality sequence start: 15
High quality sequence stop: 850.
Location/Qualifiers
1. 1040
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4850526"
/clone_lib="NIH_MGC_113"
FEATURES
source

```

```

/lab host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(C). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 231 a 265 c 220 g 324 t
ORIGIN

alignment_scores:
Quality: 1195.00 Length: 313
Ratio: 4.410 Gaps: 6
Percent similarity: 86.581 Percent identity: 81.789

alignment_block:
US-09-757-781-2 x BG745723/rev ..

Align seg 1/1 to reverse of: BG745723 from: 1 to: 1040

721 GYIleGluGlyLeuAspGluSerProSerArgAsnAlaAla...LeuSe 736
|||||:||||| :|||:|||||:|||||
995 GGGATTAAAGCCCTTGATGAATTTGCCAGCAGAAATGGCCCTCCAGC 946
736 rArgIleMetGlyLysSerGlyLysTyrGlnLeuSerProThrValAsn 753
945 TAGATTATAGGGGTACAAATCCACCCCGTCCCTCAACAGGGAATATGGCC 896
753 etProGlnAspAspThrValIleIleGluAspAspArgLeuPro.Valle 769
895 CCAAGAGGAAACAGCGTCACGTACTAGAAATGACATTGGCCAGAGTGT 846
769 uProProHisLeuSerAsp.GlnSerSerSerSerSerHisAspAspVal 785
845 TCCTCCACATCTCTGTGACACGCTCGTCCAGCTCCCATGATGATGTG 796
786 GlyPheValThrAlaAspAlaGlyThrTyrAlaLysAlaAlaIleSerAs 802
795 GGGTTGTGAGCGCAGAGTGTGACTTGGGCCAAGGCTCCATCACTGTA 746
802 pSerAlaAspCysSerLeuSerProAspValAspProValLeuAlaPheG 819
745 TTGACCGGACTGCTCTGAGTCCAGATTTGATCCAGTCTTGCTTTTC 696
819 LnaArgGluGlyPheGlyArgGln.SerMetSerGluLysArgThrLysG 835
695 AACCGAAGAGATTTGGACGTCAGAGATGTGCAGAAAACGCCACAAGCA 646
835 n.PheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSerLys 851
645 ATGTTTCAGATGCCAGTCATTTGATTTGCTTAAACACCAAAATGTAAA 596
852 SerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAspAs 868
595 AGCATGATTTAGCT..... 581
868 pGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyLeuL 885
580 .....AGTTTCGCCAGCAGAGATGTGGTCTCTCCCTGGGCTCTGA 541
885 yslYsSerSerSerLeuGluSerLeuGlnThrAlaValAlaGluValThr 901
540 AGAAGTCAGAGCTCATGAGAGCTCTCACACCGCACTTCCAGAGTGACT 491
902 LeuAsnGlyAspIleProPheHisArgProArgProArgIleAlaArgG 918
490 TTGAATGGGGAATATCTTCCATGCTCACGGCCGGATATATCAGAGG 441
918 yArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyrAspL 935
|||||

```

440 CAGGGATGCAATGAGAGCTTCAGAGCTGCCATCGACAAATCTTATGATA 391
935 ysfProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGluAsp 951
|||||
390 AACCCCGGTAGATGATGATGATGAAGGCGATGAGACCTTGGAAGAAGAC 341
952 ThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSerAs 968
|||||
340 ACAGAGAAAGTTCAAGATCAGGAGAGAGTGTGTATCCACAGCCAGTGA 291
968 pGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGluysG 965
|||||
290 TCAGCCTTCCTCCACTCTCTGAGAGACAAATGAATGGAACCAAGAGAAAG 241
985 lAspArgThrAspArgLysLysAspLysThrGlySerGluLysLysLys 1001
|||||
240 GTGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 191
1002 AspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLeu 1018
|||||
190 GATGAGATTAAGAGAGAGATTAATGAATGAAGCCAGAGAGGATGCTGAA 141
1018 sGlyLeuGlyAspMetPheArgPheGlyLys 1028
|||||
140 GGGCTTGGAGACATGTCTGAGCTTGCCCAA 110
seq_name: gb_est2: BG170486
seq_documentation_block:
LOCUS BG170486 713 bp mRNA linear EST 06-FEB-2001
DEFINITION 60232859F1 NIH_MGC_89 Homo sapiens CDNA clone IMAGE:4425975 5',
mRNA sequence.
ACCESSION BG170486
VERSION BG170486.1 GI:12677189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 713)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10172 row: e column: 16
High quality sequence scop: 710.
FEATURES
source
location/Qualifiers
1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4425975"
/clone_lib="NIH_MGC_89"
/tissue_type="hypertrophoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; site: 1: NotI;
site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 213 a 167 c 206 g 127 t
ORIGIN
alignment_scores: Quality: 1181.00 Length: 235

Ratio: 5.069 Gaps: 0
Percent Similarity: 99.149 Percent Identity: 99.149
alignment_block:
US-09-757-781-2 x BG170486 ..
Align seq 1/1 to: BG170486 from: 1 to: 713
1034 LysAlaLeuThrThrGlyLysIleValIleGlnGluSerPheThrSergl 1050
|||||
2 AAGATTAGAAACGGGTAAATTAATTAATACAGAAATCTTTACATCAGA 51
1050 uGluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysT 1067
52 AGAGAGAGAGATAGCAATGAAGAGAGAGAGAGAGATTCACACCAAAA 101
1067 hArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIle 1083
102 CTCGAGAAATTTAGGGAACGACACAGCTCGAGAGCTGATGCTGAATAT 151
1084 GlnAspPheHisArgThrPheGlyCysAspAspGluLeuMetLysGly 1100
152 CAGATTTTCATCGGACATTTGGCTGTGATGATGATGATGATGATGATG 201
1100 yValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSer 1117
202 AGTTTCTTCTATGATGAGTTCATGGCTTCACAGCTAGACTCAGAGCC 251
1117 roArgGluGlyHisMetMetAlaLeuTyrAlaGlnValLysLysPro 1133
252 CACGAGAGGGGCATATGATGATGATGATGATGATGATGATGATGATG 300
1134 ArgAsnSerLysProSerProValAspSerAsnArgSerThrProSerAs 1150
301 CGGAATTCACAAACCTCACCCTGATGACAGTAAACATCACTCCTAGCA 350
1150 nHisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGln 1167
351 TCATGATCGGATACAGCGTCTGAGGCAAGATTTACAGACCAAGCAAG 400
1167 sArgLysAspValGluAspArgArgArgThrTyrSerPheGluGlnProTyr 1183
401 ATGAAAGATGTAAGAGATGTCGCGGACCTTACTTTGACCAACCTCGG 450
1184 ProAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValG 1200
451 CCGAAGCGACGCGCGCGGACGACAGCGGCGGACACTGGTGTCCGTGA 500
1200 uValGlnMetGlnArgGlnArgGlnGluArgGluArgLysSerSerGln 1217
501 GGTGCATGATGAGCGGCGGCGGAGAGAGAGCGGAGAGCTCCAGACAG 550
1217 lArgIleArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSer 1233
551 CCCAGCCCGACGATACAGCTCTGCTCGGCAACAGAGAAAAAGCCAGC 600
1234 SerValSerGlnAspSerTyrGluGlnAlaProTyrSerProGlyGlu 1250
601 TCGGTCTCCCAAGGACTCTTGAGACAGACTACTCCCTGGGAGAGGCTT 650
1250 eGlnSerAlaLysGluAsnProArgTyrSerSerTyrGlnGlnLysArg 1267
651 CCAAGATGCCAAAGAGAAAGAACCCAGAGTACTCCAGGCTACAAAGCTC 700
1267 snGly 1268
701 ACGGT 705
seq_name: gb_est2: BG910244
seq_documentation_block:
LOCUS BG910244 694 bp mRNA linear EST 05-JUN-2001
DEFINITION 602805735F1 NCI_CGAP_Brn67 Homo sapiens CDNA clone IMAGE:4937865

sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing: 5' end primer [CTTCTCTCTTAAAGCTGCG], 3' end primer [CGACGTGACGCTGACACA].

BASE COUNT 259 a 161 c 246 g 146 t 2 others
ORIGIN

alignment_scores:
Quality: 1146.00 Length: 245
Ratio: 4.815 Gaps: 0
Percent Similarity: 97.143 Percent Identity: 88.571

alignment_block:
US-09-757-781-2 x AU079784 ..

Align seg 1/1 to: AU079784 from: 1 to: 814

```

855 LeuGlyIleAlaSpGluThrLysLeuAsnThrValAspAspGlnLysAl 871
|||||.....|.....|.....|.....|.....|.....|.....|
74 TTAACAGTAGCTGACGAGCTAACTCAATACAGTGGATGACGAGAGC 123
871 aglySerProSerArgAspValGlyProSerLeuGlyLeuLysLysSers 888
|||||.....|.....|.....|.....|.....|.....|.....|
124 AGGCTCTCCAGTAGAGATGAGGACCCCTTAGCTCGAACAATCTA 173
888 erSerLeuGluSerLeuGlnThrAlaValAlaGluValThrLeuAsnGly 904
|||||.....|.....|.....|.....|.....|.....|.....|
174 GCTCCTTAAAGTCTGACGACCGCTGTGCGGAGGTGAGCGCTGAACGG 223
905 AspIleProPheHisArgProArgProArgIleIleArgGlyArgGly 921
|||||.....|.....|.....|.....|.....|.....|.....|
224 ACATTCCTTCCACCGCCCTGCGCCACGAATCATCGGGAGGGGCTG 273
921 sAsnGluSerPheArgAlaAlaIleAspLysSerTyrAspLysProAla 938
|||||.....|.....|.....|.....|.....|.....|.....|
274 CAACGAAACCTTCAGAGCCGCCATTGACAAAGTCTTCGACAAAGCCATG 323
938 aLAspAspAspAspGluGlyMetGluThrLeuGluGluAspThrGlu 954
|||||.....|.....|.....|.....|.....|.....|.....|
324 TTGATGATGACGACGAAAGCATGAGACCTTGGAAAGAACACAGAGAA 373
955 SerSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnPro 971
|||||.....|.....|.....|.....|.....|.....|.....|
374 ACCTGAGGTGAGGAGGAGTCCGTCCACGTCCAGCGATCAGCCTTC 423
971 rHisSerLeuGluArgGlnMetAsnGluAngluLysGlyAspLys 988
|||||.....|.....|.....|.....|.....|.....|.....|
424 CTATCTCTGGAAGAGCAATGAAATGAGACCCAGAAAGGAGACAAG 473
988 hAspArgLysLysAspLysThrGlyLysGluLysLysAspArgAsp 1004
|||||.....|.....|.....|.....|.....|.....|.....|
474 CAGAGGAGGAAAAAGCAAGCCCGGAAGAGTAAGAAAGAACCCAGAG 523
1005 LysGluLysAspLysMetLysAlaLysLysGlyMetLeuLysGlu 1021
|||||.....|.....|.....|.....|.....|.....|.....|
524 AAGGAGAAAGATTAACCTGAAGCCAAAGAGGGGATCTGAAGGCTTGG 573
1021 yAspMetPheArgPheGlyLysHisArgLysAspAspLysIleGluLys 1038
|||||.....|.....|.....|.....|.....|.....|.....|
574 GGACATGTTGAGGTTGGCAACATGCAAGAAAGATGCAAGATGAAAAAA 623
1038 hGlyLysIleLysIleGlnGluSerPheThrSerGluGluGlyIle 1054
|||||.....|.....|.....|.....|.....|.....|.....|
624 TGGGTCGAATTAATAATCCAGGATCTTCACTCAAAAGAGGACAGGAG 673
1055 ArgMetLysGlnGluGlnGluArgIleGlnAlaLysThrArgGluPhe 1071
|||||.....|.....|.....|.....|.....|.....|.....|
674 CGGATCAAGAAAGAACGAGAGAGATTCAAGCAAAACTCGAGAGTTTAA 723

```

```

1071 gGluArgGlnAlaArgGluArgAspTyrAlaGluIleGlnAspPheHis 1088
|||||.....|.....|.....|.....|.....|.....|.....|
724 GGAGCGGCAAGCCGAGAGGCTGATATGACAGATTCAGATTTCCATC 773
1088 rGthrPheGlyCysAspAspGluLeuMetGly 1099
|||||.....|.....|.....|.....|.....|.....|.....|
774 GGACGTTTGGCTGTGATGACAAAGTCTGTTGGG 808

```

seq_name: gb_est1:AL529632

seq_documentation_block:

LOCUS AL529632 973 bp mRNA linear EST 13-FEB-2001
DEFINITION AL529632 LTR_NFL001_NBC4 Homo sapiens cDNA clone CS0DD005YD18 5
prime, mRNA sequence.
ACCESSION AL529632
VERSION AL529632.1 GI:12793125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 973)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE L4.W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source
1..973
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD005YD18"
/clone_11b="LTR_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 209 a 290 c 286 g 171 t 17 others
ORIGIN

alignment_scores:
Quality: 1132.00 Length: 231
Ratio: 5.122 Gaps: 1
Percent Similarity: 95.671 Percent Identity: 95.238

alignment_block:
US-09-757-781-2 x AL529632 ..

Align seg 1/1 to: AL529632 from: 1 to: 973

```

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||.....|.....|.....|.....|.....|.....|.....|
281 ATGAAGTAAACCTTKTKCTTGGACGACCGGGGTGCTCTKCCGTGCG 330
17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
|||||.....|.....|.....|.....|.....|.....|.....|
331 GGACGGCCACATGAWAKTTTTCAGCCCTCATCCAGCAGCGTGAACCCGT 380
34 yTArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||.....|.....|.....|.....|.....|.....|.....|
381 ACCGGAAGGCCATCGCCAAAGATCCAAACTAGTGATACAGGTGACATCG 430

```

```

51 LeuGIuHISGLYAspGlyIleLeuAspLeuAspIleLeuCysAs 67
|||||
431 TTGGAACTGTGAGATGAGGAACTAGACCTTGATACACTTCTTGGA 480
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluAspP 84
|||||
481 TGTAGCAGACGATAAAKACACAGCTGGTAGCAGTTTGTGAGACAGATC 530
84 roHISHisGLYAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
531 CACATTCACGAGGATKATGCGACAGTCCAGTCCAGGGGTACCCAGAGC 580
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117
581 CCAGAGATATTGTGTTAGCTTGGCACCACCAAAAGTCTCCAGCCCTTCA 630
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValIleuArg 134
631 GCCTTACCAAGCAACAAAGTGAATGAGTCAACCTTACGCTTCGAG 680
134 laAsnMetProLeuHISValArgArgSerSerAspProAlaLeuIleGly 150
681 CAAATATGCTCTCTCATGTTGACGCAATAGTTGACCCAGCTCTAATTGC 730
151 LeuSerThrSerValSerAspSerAsnPheserSerGluGluProSerAr 167
731 CTCTCCACTTCTGTCAGTATGATATTTTCTCTGAGAGCCTTCAAG 780
167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsn 184
781 GAAATATCCACAGCGCTGTCACACACAGCTGCTCTCCACAGCAGAAC 830
184 hPAla.GlySerProLysThrCysAspArgLysLysAspGluAsnTyrAr 200
831 CTGCGGGAGTCTCTAAACCTGCGACAGAGAAAGATTAACAACTACAG 880
200 gSerLeuProArgAspThrSerAsnThrPheSerAsnGlnPheGlnArgAsp 217
881 AAGCTCCCGGGGATGACTGATGATGCTGCTAAACCAATTTCACAGAGACA 930
217 snAlaArgSerSerLeuSerAlaSerHisProMetValGly 230
931 ATGCTCGCTGCTCTGAGTGCACAGTCCACCAATGTTGTTGGC 971

seq_name: gb_est2:B1464081
seq_documentation_block:
LOCUS B1464081 830 bp mRNA linear EST 21-ANG-2001
DEFINITION 603202858f1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:526846 5',
ACCESSION B1464081
VERSION B1464081.1 GI:15254737
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 830)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraiki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
http://image.llnl.gov
Plate: LLM11677 row: d column: 15

```

```

FEATURES             High quality sequence stop: 722.
source               Location/Qualifiers
                    1..830
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_image="526846"
                    /clone_id="NIH_MGC_97"
                    /lab_host="DH10B"
                    /note="Organ: testis; Vector: pBluescript (modified
                    pBluescript KS+); Site_1: BamHI; Site_2: SalI; XhoI (gtcgag
                    ); Oligo-dn primed using primer 5'-TTTTTTTTTTTTTTTAA-3',
                    size-selected for average insert size 2.2 kb and
                    normalized to 10^5. This is a primary library enriched
                    for full-length clones and constructed using the
                    Cap-trapped method (Carninci, in preparation). Library
                    constructed by M. Brownstein (NIH/NHGRI, National
                    Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT          247 a 191 c 204 g 188 t
ORIGIN
alignment_scores:
  Quality: 1127.50      Length: 315
  Ratio: 4.528          Gaps: 7
  Percent Similarity: 79.048  Percent Identity: 74.603
alignment_block:
US-09-757-781-2 x B1464081 ..
Align seq 1/1 to: B1464081 from: 1 to: 830
132 LeuArgAlaAsnMetProLeuHISValArgArgSerSerAspProAlaLe 148
::: |||||
21 ATGACACGCTGATATGCTCTTCAAGTTTCGACGACGATGAGACCCAGCTCT 70
148 uIleGlyLeuSerThrSerValSerAspSerAsnPheserSerGluGluP 165
||| |||||
71 AATTGGCTCTCCACTTCTGTCAGTATGATATTTTCTCTGAGAGAC 120
165 roSerArgLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLys 181
||| |||||
121 CTTCAGAGAAATATCCACACAGCTGTCACAAACAGCTGGCTCTCTCAAG 170
182 GlnAsnThrAlaGlySerProLysThrCysAspArgLysLysAspGluAs 198
||| |||||
171 CAGAACACTGCTGGGAGTCTTAACCTGCCACAGAGAG..... 209
198 nTyrArgSerLeuProArgAspThrSerAsnThrPheSerAsnGlnPheGln 215
209 .....
215 rgAspAsnAlaArgSerSerLeuSerAlaSerHisProMetValGlyLys 231
209 .....
232 TrpLeuGluLysGlnGluGlnAspGluAspGlyThrGluLysAspAsnSe 248
||| |||||
210 ..... GATGAGAGTGGGACAGCAACAGAGTAAACAG 238
248 ArgValGluProValGlnHisAlaAspThrGlyLeuGluHisIleProca 265
||| |||||
239 TCGGTTGGAACCTGTGACATGCTGACAGGGTTTGGAGCAATATGCCCA 288
265 snPheSerLeuAspAspMetValLysLeuValGluValProAsnAspGly 281
||| |||||
289 ACTTTTCTGTGAGATATGTAAGCTGTAGAACTCCCAACAGATGA 338
282 gLysProLeuGlyIleHisValAlaProPheSerAlaArgGlyLysArgTh 298
||| |||||
339 GGGCGCTGTGGGAATTCATGATGCTTTCAGTGTCTCGAGCGCGCAGAAC 388
298 rLeuGlyLeuLeuValLysArgLeuGluLysGlyLysAlaGluHisG 315
|||||

```

389 CCTGGGTTATTAGTAAACGATTCGAGAAAGTGGTAAAGCTGAACATG 438
315 luasnleupheargluasnaspcysilevalargilleasnaspglyasp 331
439 AAAATCTTTTTCGTGAGAAATGATTCATGTCAGATTAATGATGGGAC 488
332 leuargasnargargphegluginalaglnhismetphearaglinalame 348
489 CTTCGAAAGAAGATTGTGAAACAGCAACATATGTTCCGCAACCAT 538
348 lartgthrproillelterpphenisvalvalproalalaasnlysgl 365
539 GCGTACACCATATTTGTTCCATGTGTTCTCGCAGCAATTAAGAGC 588
365 lntyriglunleuserglinserglulysasn. Tyrtyrsersear 381
589 AGTATGACACATATCCCAAGAGAGACAAATTACTATTCAAGCCG 638
381 gpheserproaspsergln...Tyrilleasnasnargservalasnsera 397
639 TTTTACCCCTGCACGAGTATGATGACCAACGAGGTGACACAGT 668
397 la...glyleuhis Thrvalglinarala..Proargleuasnhspr 411
689 GCAGGGCTTCCACAGGGTGCAGAAAGACCCCGATGCAACAAACC 738
411 oproguglinleaspsr.Hlaserargleuprophisserialahsapro 427
739 GCGTAGACGATGACTCTCTCATCAAGAAATACATAGAGGAAAGGCC 788
428 Serglylyluproseralaproalaseralaproglin 440
789 CTCGGGAAACCCCATCGGTCCACTCGGACCCAAAG 827
seq_name: gb_est2:BG336977
seq_documentation_block:
LOCUS BG336977 941 bp mRNA linear EST 27-FEB-2001
DEFINITION 602434429P1 NIH_MGC_46 Homo sapiens cDNA IMAGE:4552423 5',
mRNA sequence.
ACCESSION BG336977
VERSION BG336977.1 GI:13143415
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 941)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.lnl.gov
plate: LICM1248 row: 1 column: 08
High quality sequence stop: 638.
FEATURES
source
1..941
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4552423"
/clone_lib="NIH_MGC_46"
/tissue_type="lelomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site.1: XhoI; Site.2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:"

GCACAGC(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the Laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

BASE COUNT 280 a 228 c 236 g 197 t
ORIGIN

alignment_scores:
Quality: 1118.50 Length: 302
Ratio: 4.285 Gaps: 9
Percent Similarity: 86.424 Percent Identity: 80.795

alignment_block:
US-09-757-781-2 x BG336977 ..

Align seg 1/1 to: BG336977 from: 1 to: 941

218 Alaargserleuseralaserhisprometvalglyttrpleugl 234
1 GCTGCTCGTCTCTGAGTGGCCAGTCCCAATGTTGGCAAGTGGTGA 50
234 ulysglngluglnaspglylthrglugluasnasnserargvalg 251
51 GAAACAGAACAGATGAGATGGAGACAGAAAGAGATACAGTGTGTG 100
251 luprovalglynhisAlasprhnglyleughislietproaspheser 267
101 AACCTGTGGACATGCTGACACGGGTTTGGAGCATATACCAACTTTCT 150
268 leuaspasphetvallysleuvalgluvalproasnaspglylprole 284
151 CTGGATGATATGTTAAAGCTGTAGAAAGTCCCAACATGAGGGCTCT 200
284 uglylthhisvalvalpropheserlathrglylathrglyl 301
201 GGGAAATCATATGTAAGTCTTCAAGTCTCGAGGGGAGAAACCTGGGCT 250
301 euleuvallysargleughlysglyllysAlagluhishgluasleu 317
251 TATTAGTAAACGATGTGAGAAAGTGTAAAGCTGAACATGAATACTTT 300
318 pheargglu.Asnaspcysilevalargilleasnaspglyaspleuarg 334
301 TTTCGTAGACATGATGTCATTTAGATTAATGATGGCACCTTCGAA 350
334 snarg.Argphegluglnalaglnhismetphearaglinalametarqth 350
351 ATAGAACGATTTGAAACAGCACACATATATGTTGCCAAGCCATGCGTAC 400
350 rproillelterpphenisvalvalproalalaasnlysglunlntyr 367
401 ACCCATATTTGTTCCATGTGTTCTCGCAGCAATTAAGAGAGATG 450
367 luglnleuserglinserglulysasnantyrtyrsersearpheser 383
451 AACGACATATCCCAAGAGAGAAACAAATTACTTTCAAGCGGTTTACG 500
384 proaspserglntyrilleasnasnargservalasnseralaglileu 400
501 CTGACAGACCCGTATTTGACACAGAGTGTGAACAGTGCAGGGCTTCA 550
400 sthvalglathrglathrgleuasnhsprproglunlntleasps 417
551 CACGTCACAGACACCCGACTGAACACACCCCT.GAGGAGATGAGT 599
417 ernhisserargleuprohisserialahs.Proserglyltyrproprose 433
600 CTCACCTCAAGACTACGCAT...AGGCAACCCCTCGGGAATAAGCATC 646
433 rAlaproalaseralaprogluasnvalpheserThthrvaliserSerg 450

```
|||||
647 GGCTCCAGCCTCGCACCTCAGCAATGTATAGTACGACTGTAAAGACGT 696
450 LTTyrAsnThrLysIleGlyLysArgLeuAsnIleGlnLeuLysLys 466
697 GGTAAATACACCAACAATAATAGCCAGAGGTAAATATCCACTTCAGAAA 746
467 GlyThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleG1 483
747 GGGACCGAAGGCTGGGAT...CAACCTACTTCCAGA...ACATGGTACCA 790
483 yGlySerAlaProIleTyrValIysAsnIleLeuProArgGlyAla.... 498
791 AGGTGCTGCTCCATCGAGTCAAAA...CTCTCCCCGGTGGGCCAAGAA 837
499 .....AlaIle.GlnAspGlyArgLeuLysAlaGlyAspArgLeuIle 512
838 GGATGCCGNTAAAGACGAGGAGGAGACTCCAGCGCGGTCTGTATGCTT 865
```



```

4445 AGGTCAAGCCCGAAGAAAGTGAAGAAACACTCAGCTCAAGACTAC 4494
575 sp1leuValleuthrProAspGlyThrArgLupheleuthrPhegluVal 591
4495 GCTTGTTC.....ACTGAAAGAAATACATTGAGGTA 4526
592 ProLeu...AsnAspSerGlySerAlaGlyLeuGlyValSerValysgl 607
4527 AAATTTATTTAAATAGTCAGGTCTAGATTCAGTTTCTTCGAGAGA 4576
607 ysnAspSerGlyuAsnHisAlaSpluGlyllePheValysSerI 624
4577 TAACTTTATACCGGAGCAATTAATCCAGCATAGAGGTTAAAAAGC 4626
624 lelleAsnGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAsp 640
4627 TCTTCTCTGACAGCCAGCAGCAAGAGTGAATAATTGATGTAGAGAT 4676
641 GlnLeuIleAlaValAsnGlyLysLeuLeuGlyLysThrAsnGlnAs 657
4677 GTTATCTTGAAGTGAATGAGCCTCTTGAAGGACTATCTCAGCAGGA 4726
657 palMetGluThrLeuArgArgSerMetSerThrGluGlyAsnLysArg 674
4727 AGTCATATCTCTCTCAGGGAAGTCTCCAGAA..... 4760
674 lymetIleGlnLeuIleValAlaArgArgIleSerLysCysAsnGluLeu 690
4761 ....GTATCTTGTCTCTCTGCAGA..... 4781
691 LysSerProGlySerProGlyProGlyLeuProIleGluThrAlaLe 707
4782 .....CCTCACCCTGGTGTACCGGAATATGATCTGCGGT 4819
707 uAspArgGluArgArgIleSerHisLeuTyrSerGlyIleGlu 724
4820 T.....TTGACCCCACTTCAGTCTCCAGCAACG 4848
724 lyleuAspGlySerProSerArgAsnAlaAlaLeuSerArgIle..... 738
4849 TACTTCCAAACAGCAGTAAGACTCTTCTCAGCCATCATGTGTGAGCAA 4898
739 .....MetGlyLysSerGlyLysTyrGluLe 747
4899 AGCACCAGCTCAGATGAATGAATGATCAGACAAAGCAAAACAGTG 4948
747 u...SerProThrValAsnMetProGlnAspArgThrValIleIleGlu 763
4949 CAAGTCCCAATCCAGAGAGACAGTTCAGTACAGCAGTGGAGTGAG 4998
763 spAspArgLeuProValLeuProHisLeuSerAspGlnSer...Ser 778
4999 AAGATGACTTAGTCACAGCTCCAGCAACATATCAATTCGACCTGAGT 5048
779 SerSerSerHisAspAspValGlyPheValThrAlaAspAlaGlyThr 795
5049 TCAGCTTTCAGACTCTAAGCAACATGATCAGACAGCAGCAGAGTCA 5098
795 palAluysAlaIleSer..... 801
5099 TCATGTAGACACCAGAGTCAAGAGATACCATTTGTACCATGTTTACT 5148
802 .....AspSerAlaAspCys 806
5149 ATCTTCAGAAATATCCCAATTAACAGAGTTTGAGAGCAGACATATCCTCC 5198
807 SerLeuSerProAspValAspProValIleuAlaPheGlnArgGluGly 823
5199 CCTCTACCAACCGGATATGGCTCTGGGCAAGATTAACAACCCCAATCAGA 5248
823 eGlyArgGlnSerMetSer 829
::: ::::: |||

```

```

5249 ATCTGCTTCTCTAGTTCC 5267
seq_name: /cgn2_6/plodata/2/lna/5A_COMB.seq:us-08-596-291-1
seq_documentation_block:
; Sequence 1, Application US/08596291
; Patent No. 5821075
; GENERAL INFORMATION:
; APPLICANT: GONZ, LEONEL JORGE
; APPLICANT: SARAS, JAN
; APPLICANT: CLAESON-WELSH, LENA
; APPLICANT: HELDIN, CARL-HENRIK
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596,291
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: 10461/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; TELEX: 92-1742 EZEKIEL
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8040 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..7475
; US-08-596-291-1

alignment_scores:
Quality: 327.50 Length: 927
Ratio: 0.741 Gaps: 34
Percent Similarity: 47.681 Percent Identity: 20.928

alignment_block:
US-09-757-781-2 x US-08-596-291-1 ..
Align seg 1/1 to: US-08-596-291-1 from: 1 to: 8040
||| |||||
80 AspGlnGlnAspProHisHisGlyGlyAspGlyThrSerAlaSerThr 96
2712 GATGCCCAAGATTAAT.....GAGAGAGCTTGTGTTAGAGACCT 2749

```



```

4430 TCAGATGCCCAAGTCAAGCCCAAGAAAAAGTGAAGAAAAACAATCAGC 4479
571 LAGLUSPGLUASPGLIValleuThrProaspGlyThrArgGluPheLeu 587
4480 TCAAGACTACAGCTTTC.....ACTGAAGAAAT 4511
588 ThrPheGluValProLeu...AsnAspSerGlySerAlaGlyLeuGlyVal 603
4512 ACATTGAGGTAAATATTATAAATAGCTAGCTAGATTAGTT 4561
603 LserValIlySGlyAsnArgSerIlyGlnHisAlaAspLeuGlyIleP 620
4562 TTCTCAGAGATTAATCTTATACCGAGCAAAATTAATGCCAGCATGTA 4611
620 heValIlySerIleIleAsnGlyValAlaIleSerIlyAspGlyArgLeu 636
4612 GGGTTAAAAAGCTTCTTCTGCGACACGACGACGAGAAAGTGAATAAT 4661
637 ArgValAsnAspGluLeuIleAlaValAsnGlyIlySerLeuLeuGly 653
4662 GATGTAGGAGATGTATCTTGAAGAATGAGAGCTCTTGAAGAGACT 4711
653 sThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSerThrGlu 670
4712 ATCTCAGCAGGAAGTCATATCTGCTCAGGGGAGACTGTCAGAA... 4757
670 IlyAsnIlyArgIlyMetIleGluLeuIleValAlaIleArgIleSerIy 686
4758 .....GTAATCTTCTCTCTGTCAGA..... 4778
687 CysAsnGluLeuIlySerProGlySerProGlyProGluLeuProI 703
4779 .....CCTCCACCTGGTGCTGCTACCGGAAT 4804
703 eGluThrAlaLeuAspAspArgGluArgIleSerHisSerLeuTyr 720
4805 TGAATCTGCGCTT.....TTGACCCCACTTCAGT 4833
720 eGlyIlyLeuGlyLeuAspGluSerProSerArgAsnAlaIleLeuSer 736
4834 CTCACACACAAGTACTTCCAAACAGCAATGTAAGACTTCTTCAGCCATCA 4883
737 ArgIle.....MetGlyIlySerG 743
4884 TGTGTGAGACAAGACACAGCTCATGTAAGAAATGTCAGACAAAG 4933
743 IlyIlyTyrGlnLeu...SerProThrValAsnMetProGlnAspAspThr 759
4934 CAAAAACAGTGCACACTCCCATCCAGAGACAGACTTACAGTACAGCA 4983
759 alleIleGluAspAspArgLeuProValLeuProHisIleuSerAsp 775
4984 GTGGAGGTGAGAGATGACTTGTACACGCTCCAGCAACATATCAAT 5033
776 GlnSer...SerSerSerSerHisAspAspValGlyPheValThrAla 791
5034 TCGACCTGGAGTTCAGCTTGTACAGACTCTAAGCAACATGTTATCACA 5083
791 PalagIlyThrTPAlaIlyAlaIleSer..... 801
5084 GGCACAGAGTCATCATGAAGCACCCAAAGTCAAGAAATACCATTTGTA 5133
802 .....Asp 802
5134 CCATGTTTACTCTCTCAGAAAAATCCCAATAACACAGGTTTGAGGAC 5183
803 SerAlaAspCysSerLeuSerProAspValAspProValLeuAlaPheG 819
5184 AGTAATCTCTCCCTTACACACCGGATATGCTGGGAGAGATTATCA 5233
819 natGglugIlyPheGlyArgGlnSerMetSer 829

```

5234 ACCCCAAATCAGAAATCTGCTCTCTACTGCG 5264

seq_name: /cgn2_6/plodata/2/lna/6A_COMB.seq:us-09-100-804-1

seq_documentation_block:

Sequence 1, Application US/09100804

Patent No. 6066472

GENERAL INFORMATION:

APPLICANT: GONZ, LEONEL JORGE

APPLICANT: SARAS, JAN

APPLICANT: CLAESON-WELSH, LENA

TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/100,804

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/596,291

FILING DATE: 09-AUG-1996

APPLICATION NUMBER: US 08/115,573

FILING DATE: 01-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/09943

FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

REFERENCE/DOCKET NUMBER: 10461/7003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 8040 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: HOMO SAPIENS

FEATURE:

NAME/KEY: CDS

LOCATION: 78..7475

US-09-100-804-1

alignment_scores:

Quality: 327.50 Length: 927

Ratio: 0.741 Gaps: 34

Percent Similarity: 47.681 Percent Identity: 20.928

alignment_block:

US-09-757-781-2 x US-09-100-804-1 ..

Align seg 1/1 to: US-09-100-804-1 from: 1 to: 8040

```

80 AspgIngluAspProHISHisGlyAspGlyThrSerAlaSerSerth 96
   ||| |||||
2712 GATGCCCAAGATATTT.....GACAGACCTCGTTTAGAGACCT 2749
   ||| |||||
96 rGlyThrGlnSerProGlnIlePheGlySerGluLeuGly..... 109
   ||| |||||
2750 GAATCTCCACAGACAGACTCTGTAGAGATTAATATGGACGACATCA 2799
   ||| |||||
110 .....ThrAsnValSerAlaPheGlnPro 118
   ||| |||||
2800 GCACGTGCAGCTGGCCAGACGACACCTCCACAACTCTGCTGTGACCT 2849
   ||| |||||
119 TyrGlnAlaThrSerGluIle.....GluValThrProSerValLeuAr 133
   ||| |||||
2850 TTATGATGTTCAAGCTGAGATTTGAAAGAGCTTCCTGCTCAGACCTGC 2899
   ||| |||||
133 galaSmetProLeuHis.....ValArgArgSerSerSerProAlaL 148
   ||| |||||
2900 GCTTACCGACCATTCGAAACAGTTCAAAAGACAAAGATGACAAAGCTT 2949
   ||| |||||
148 eu.....IleGlyLeuSerThrSerValSerAspSerAsp 159
   ||| |||||
2950 CARGGAGGAAAAAGCCTAGAGAGATGATGATTAATCATACATGATCTCAGT 2999
   ||| |||||
160 PheSerSerGluGluProSerArgLysAsnProThrArgThrSerThrTh 176
   ||| |||||
3000 CAGGCCCTCTCTATCCACATCGGAAAAAT..... 3029
   ||| |||||
176 rAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCyAspA 193
   ||| |||||
3030 .....GTCAATTGTTAACATGAGAACCCCAACCAACCCGTTGCGAG 3069
   ||| |||||
193 rGlyLysAspGluAsnLysArgSerLeuProArg...AspThrSerSer 208
   ||| |||||
3070 AGTTGGTGGGAAACCTTCTCCACAGATGTCAAGATCTGATGACGAATCT 3119
   ||| |||||
209 TrpSerSerGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSe 225
   ||| |||||
3120 TTGGCAGAGACTGCACAACTTAATTAATCAAAAGTCTGTGCGACTTTAAA 3169
   ||| |||||
225 rHisPrometValGlyLysTrpLeuGluLysGlnGluGlnAspGluAsp 242
   ||| |||||
3170 T.....AGAGTCTGAAAGAGAGCA 3189
   ||| |||||
242 LysThrGluGluAspAsnSerArgValGluProValGlyHisAlaSerThr 258
   ||| |||||
3190 AACATGAATCAGACTCTCCATCATTTGAGACCCCTGGGCAAGCA..... 3233
   ||| |||||
259 GlyLeuGluHisIleProAsnPheSerLeuAspAspMet..... 271
   ||| |||||
3234 .....TATGTTCTAGATGTCCTACACAAAAAGATG 3262
   ||| |||||
272 .....ValLysLeuValGluValProA 279
   ||| |||||
3263 GAGCATATGATCTTACACAGAAAGAGATCTCTACTGAAACCTGAA..... 3312
   ||| |||||
279 snaspglyGly...ProLeuGlyIleHisValAlaProPheSerAlaArg 294
   ||| |||||
3313 AAGATGCAAAAGTATGGCTTGGATTTCAAATTT..... 3347
   ||| |||||
295 GlyGlyArg.....ThrLeuGlyLeuLeuValLysArgLeuG 307
   ||| |||||
3348 GGTGGGGAAGAGATGAGACTGACCTTGGCATTTTATTCAGCTCAGTTGCG 3397
   ||| |||||
307 LysGlyLysLysAlaGluHisGluAsnLeuPheArgGluAsnAspCysI 324
   ||| |||||
3398 CCGTGGAGACCACTGCTTCATGATGCTTTGAAAGCCAGAGACCCCTT 3447
   ||| |||||
324 LevalArgLysAspGlyAspLeuArgAsnArgArgPheGlnGluAla 340
   ||| |||||
3448 TGATATCTGTGAATAGTGTAGTGTGAGAGGAGTGCACGACCATGCTGCA 3497

```

```

341 GlnHisMetPheArgGlnAla..... 347
   ||| |||||
3498 ATTGAAATTTTGCAGAAATGCAACCTGAAGATGACACTTTTATCTGTCA 3547
   ||| |||||
348 .....MetArgThrProIleIleThrPheHisVal.... 357
   ||| |||||
3548 GCCAAAGAAAAAGATATCCAAAGTGCCTTACTCTGTGATCTACCA 3597
   ||| |||||
357 ..... 357
   ||| |||||
3598 ATGAGATGAAAAACTACATGACAAATCTCTACATGACAGACAGTCT 3647
   ||| |||||
358 ValProAlaAlaAsnLysGluGlyThr..... 366
   ||| |||||
3648 ATAGATTCTCTTCCAGAGATGATCACACTGTGACGTGTACCTGAGCA 3697
   ||| |||||
366 ..... 366
   ||| |||||
3698 CACTGTGAGAACTCTTTGGGCGCTGTGGGCGCTGCGGGAAGAGCC 3747
   ||| |||||
367 .....GluGlnLeuSerGlnSerGlu 373
   ||| |||||
3748 TGAGTCTCAAGATTTCCAGACTGAGAGTGCAGCTGTGCTCAAGCCAG 3797
   ||| |||||
374 LysAsnAsnLysThrSerSerArgPhe..... 382
   ||| |||||
3798 GTCAATGGTTCTTCTTGGCCAGCCATTAGGTGACCAACCTGCGACGATC 3847
   ||| |||||
383 .....SerProAspSerGlnLysThrLeuAspAsnArgSerValAsn 396
   ||| |||||
3848 ACAGCATGCGACGCCCTTCCCATCTGTATATCCAAACCCAGGAGAA 3897
   ||| |||||
396 eAlaGlyLeuHisThrValGln.....ArgAlaProArgLeuAsn 409
   ||| |||||
3898 AGACTTTCAGTATGATTAACCAACCAAACTTAAACCCAGGCACTTCT 3947
   ||| |||||
410 HisProGluGlnIleAsp...SerHisSerArgLeuProHisSerAl 425
   ||| |||||
3948 GATGTACTGATTAATCTCAGACCTGAGATTCAGACATGATGAGAGCC 3997
   ||| |||||
425 HisProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsn 442
   ||| |||||
3998 TTACTCCAGCAGTCAAGATCATCAACACCAACAGAGAACTTCTCTT 4047
   ||| |||||
442 AlPheSerThrThrValSerSerGlyLysAsnThr..... 453
   ||| |||||
4048 CAGTGAATACATCCAAACAAAGATGAATTTAAACCTTTCTTCATCACCT 4097
   ||| |||||
454 LysLysLysLysLysArgLeuAsnIleGlnLeuLysLysGlyThrGlu 470
   ||| |||||
4098 CCTAGGCTGAGATATCTTGTAGCTTACACTGCTTAAATATGATAACAG 4147
   ||| |||||
470 yLeuGlyPheSerIleThr.....SerArgAspValThrIleGlyG 484
   ||| |||||
4148 CTTGGGAGTAACTGTACAGGAGGTGTGAATACGAGTGTGACATGAGT 4197
   ||| |||||
484 LysSerAlaProIleLysValLysAsnIleLeuProArgGlyAlaIle 500
   ||| |||||
4198 GC.....ATTATGTGAAGATGTTATTTCCCAAGGAGACACACAG 4238
   ||| |||||
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValLysGlyVal 517
   ||| |||||
4239 TCTGATGTAGAAATTCACAAAGTGTATGCGCTGCTGCTGCTGCTGCTG 4288
   ||| |||||
517 LAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 534
   ||| |||||
4289 TAGTCTGAAGAGAGCCCATGATGACCAAGCTGTGAAACATGAGAAATA 4338
   ||| |||||
534 hTrpMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
   ||| |||||
4339 CA.....GGACAGGTGGTTCATCTGTATTGAAAAAGGACAACTCT 4379

```

```

551 PheHisProArgLuu.....LeuAsnAl 558
    ::::|
4380 CCACATCTTAAGAACATGTCCGGTAACCCACAGCTGATACCTTCAC 4429
558 agLuproSerGlnMetGlnLlePro.....LysGluThrLysA 571
    ::::|
4430 TCAGAAATGCCCAAGCTCAAGGCCGAGAAAAGTGAAGAAAACACATCAG 4479
571 laGluAspLysAspLysAlValLeuThrProAspLysThrArgLupheLeu 587
    ::::|
4480 TCAAGACCTACAGCTTGTG.....ACTGAAGAAAT 4511
588 ThrPheGluValProLeu..AsnAspSerGlySerAlGlyLeuGlyA 603
    ::::|
4512 ACATTTGAGGTAAATATTATTAATAATACCTCAGGTCTAGATTCAGTTT 4561
603 lserValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuGlyLle 620
    ::::|
4562 TTCTCGAGAAATATCTTATACCGGAGCAAAATTAAATGCCAGCATGTAA 4611
620 heValLysSerLleLleAsnGlyLysAlaLysSerLysAspGlyArgLeu 636
    ::::|
4612 GGGTTAAAAAGCTTGTGTCGACACGACGACGACGACGACGACGACGAC 4661
637 ArgValAsnAspGlnLeuLleAlaValAsnGlyLysLeuLeuGlyLys 653
    ::::|
4662 GATGTGGAGATGTATCTTGAAGGAAATGAGGCTCTTTGAAGAGACT 4711
653 srThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSerThrGlu 670
    ::::|
4712 ATCTCGACGAGGAGTCAATCTGCTCTCAGGGGACTGCTCCAGAA... 4757
670 lYAsnLysArgLysMetLleGlnLeuLleValAlaArgArgLysSerLys 686
    ::::|
4758 .....GTAATCTGCTCTCTCGACA..... 4778
687 CysAsnGlnLeuLysSerProGlySerProGlyProGlyLeuProI 703
    ::::|
4779 .....CCTCCACCTGTGTGCTACCGGAAAT 4804
703 eGluThrAlaLeuAspAspArgGluArgArgLysSerHisSerLeuTys 720
    ::::|
4805 TGATACTGCGCTT.....TTGACCCCACTTCAGT 4833
720 eArgLysLleGlnLysLeuAspGlnSerProSerArgAsnAlaAlaLeuSer 736
    ::::|
4834 CTCACGACAAAGTACTTCCAAACAGCAAGTAAGACTCTCTCGACATCA 4883
737 ArgLle.....MetGlyLysSerG 743
    ::::|
4884 TGTGTGGAGCAAAAGCACACGCTCAGATGAATAATGTACAGACAAAG 4933
743 lYsArgLysGlnLeu..SerProThrValAsnMetProGlnAspAspThy 759
    ::::|
4934 CAAAAAAGTCGACAGTCCCATCCAGAAAGACAGTTCACAGGACAGCA 4983
759 alLleLleGlnAspAspArgLeuProValLeuProProHisLeuSerAsp 775
    ::::|
4984 GTGGGAGTGAGAGATGACTTAGTCACAGCTCCACAAACATATCAAT 5033
776 GlnSer..SerSerSerHisAspAspValGlyPheValThrAlaAs 791
    ::::|
5034 TCGACTGTGAGTTCAGCTTGTGATTCAGACTCTTAAGCAACATGCTATCACA 5083
791 pAlaGlyThrTrpAlaLysAlaAlaLleSer..... 801
    ::::|
5084 GGCACAGAGTCATCATGAAGCACCACAGAGTCAAGAGATACCATTTGTA 5133
802 .....Asp 802
    ::::|
5134 CCATGTTTACTATCTCAGAAAATCCCAATTAACAGAGTTTGAGGAC 5183
803 SerAlaAspCysSerLeuSerProAspValAspProValLeuAlaPheG 819

```

```

||| ||| |||||: ||| ::::|
5184 AGTAATCCTTCCCTCTACCAACCGGATATGCTCTCGGACGAGATTATCA 5233
819 naRGluGlyPheGlyArgGlnSerMetSer 829
| ::::|
5234 ACCCCATCAGAAATCTGCTCTCTCTAGTTTG 5264

```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-290-640-45

seq_documentation_block:

```

; Sequence 45, Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Marcussen, Eric G.
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
; FILE REFERENCE: ISPH-0351
; CURRENT APPLICATION NUMBER: US/09/290,640
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 8119
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(7521)
; PUBLICATION INFORMATION:
; JOURNAL: FEBS Lett.
; VOLUME: 337
; ISSUE: 2
; PAGES: 200-206
; DATE: 1994-01-10
; DATABASE ACCESSION NUMBER: D21209/Genbank
; DATABASE ENTRY DATE: 1999-02-05
; US-09-290-640-45

```

```

alignment_scores:
  Quality: 324.50      Length: 933
  Ratio: 0.729        Gaps: 32
  Percent Similarity: 47.696   Percent Identity: 20.686

```

alignment_block:

US-09-757-781-2 x US-09-290-640-45 ..

Align seg 1/1 to: US-09-290-640-45 from: 1 to: 8119

```

80 AspGluGlnAspProHisHisGlyLysGlyAspGlyThrSerAlaSerSerTh 96
    ||| |||||
2698 GATGCCCAAGATAT.....GAGAGAGCTCGTTTGAAGACCT 2735
96 rGlyThrGlnSerProGluLlePheGlySerGluLeuGly..... 109
    ::::|
2736 GAATCTCCAGCAGAGAGTCTGTAGAGGATTAATATGAGCAGCAATCA 2785
110 .....ThrAsnAsnValSerAlaPheGlnPro 118
    ::::|
2786 GCACTGCGACTGTGGCCAGCAGCACCCTCAACAACTTGCTGTGACCT 2835
119 TyrGlnAlaThrSerGluLle.....GluValThrProSerValLeuArg 133
    ::::|
2836 TTATCAGTTCAAGCTGAGATTCTGAAGAGGCTATCTCTCAGAGCTGTC 2885
133 gAlaAsnMetProLeuHis.....ValArgArgSerSerAspProAla 148
    ::::|
2886 GCTTACACAGCATTCGCAAAAACAGTTCAAAAAGAGAAATGACAAAGCTT 2935
148 eu.....llegLysSerThrSerValSerValSerAspSerAsn 159
    ::::|
2936 CATGGAGGAAAAAGCCTTAGAGAGATGATGAATCATCATGATCTTCAGT 2985

```

160 PheSerSerGluGluProSerArgLysAsnProThrArgTrpSerThrTh 176
 : : : : :
 2986 CAGGCTCTCTCTATCCACATCGAATAAT..... 3015
 176 rAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAspA 193
 : : : : :
 3016GTCATGTGTAACTAGTGAACCCACCACCAACACCGTTGAG 3055
 193 rGlyLysAspGluAsnTrpArgSerLeuProArg...AspThrSerAsn 208
 : : : : :
 3056 AGTTGGTGGGAAACCTTCTCACAGATGTCAAGATCTGATGACGAATCT 3105
 209 TrpSerAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSe 225
 : : : : :
 3106 TTGGCAGGAGTGCAAAACTTAATTAATTCAAAGTCTGTCCGAGTTTAA 3155
 225 rHisProMetValGlyLysTrpLeuGluLysGlnGluAspGluAspG 242
 : : : : :
 3156 T.....AGAACTCTGGAAGAGAGA 3175
 242 LyrThrGluLysAspAsnSerArgValGluProValGlyHisAla..... 256
 : : : : :
 3176 AACATGAATCAGACTCTCTCATCCATGGAACCCCTGGGCAACATATGTT 3225
 257AspThrGlyLeuGlnHisIleProAs 265
 3226 CTAGAAATGACTATGCATAGTACTTGGAAACTCTTCATCCCAAGTACCTT 3275
 265 nPheSerLeuAspAspMet..... 271
 3276 AAAAGAAATGATGTCTACACAAAGATGGAGCTAGTATCTTCACCAG 3325
 272ValLysLeuValGluValIProAsnAspGlyGly...ProLeu 284
 3326 AAAGGAGATCACCCTTAGTGAACCTGAAAAAGATGCAAGTATGCGTTG 3375
 285 GlyIleHisValValIProPheSerAlaArgGlyLysArgThrLeuGlyLe 301
 : : : : :
 3376 GGAATTTCAATATATGTGGGAGAAAGATGGAAAGCTGAGTACATGAT 3425
 301 uLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLeuP 318
 : : : : :
 3426 ATTATTCAGTTCAGTTCGCCCTGAGGAGCAGCTGACTGATGATGCTGT 3475
 318 heaArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsn 334
 : : : : :
 3476 TGAAGCCAGAGAGACCGTTGTGATATCTGTGAATAGTGTGAGTGGAGGA 3525
 335 ArgArgPheGluGlnAlaGlnHisMetPheArgGlnAla..... 347
 3526 GTCAGCCACCATGCTGCATATGAAATTTTGGCAAAATGCACCTGAAGATGT 3575
 348MetArgThrProIle 353
 3576 GAACACTGTTATCTCTCAGCCAAAGAAAGAAATATCCAAAGTGCCTTCTA 3625
 353 LeTrpPheHisVal..... 357
 3626 CTCCTGTGCATCTCACAATGATGAAAACTCATGAAGAAATCTTCC 3675
 358ValProAlaAlaAsnLysGluGlnLys 366
 3676 TACATGCAAGACAGTCTATAGATTCTTCTCCACAGATCACCACTGGTTC 3725
 366 366
 3726 ACGTGTACCTGAGGACATCTGGAGAACTCTTGGGCCATCTGGGG 3775
 367Glu 367
 3776 GCCGCGCGGAAGAGCGCTGAGTTCTCAAGATTCAGAGCTAGAGTCCC 3825
 368 GlnLeuSerGlnSerGluLysAsnAsnTrpLysSerArgPhe..... 382

3826 AGCTTGCTCAAAAGCCAGGTCATATGTTCTTTGGCAGCATTTAGTGA 3875
 : : : : :
 383SerProAspSerGlnTrpIleA 390
 3876 CCAAACTGGCAGGATACACAGCATGGCAGCCCTTCCCATCTGTAAAT 3925
 390 sPAsnArgSerValAsnSerAlaGlyLeuHisThrValGln..... 403
 3926 CCAAAAGCCAGGAGAAAGAGACTTCTTACTGATATGATACCAAGCAAACT 3975
 404 ArgAlaProArgLeuAsnHisProProGluGlnIleAsp...SerHisLe 419
 : : : : :
 3976 AAAAGCCAGGCAATTTCTGATGTAACTGATTAATCTACAGACCGTGAATTC 4025
 419 rArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaProA 436
 : : : : :
 4026 AGACATGATGAAAGCCACTTCTCCAGCATGACAGATCATCAACACCA 4075
 436 LaSerAlaProGlnAsnValPheSerThrValSerSerGlyTrpAsn 452
 : : : : :
 4076 AACAGGAATCTTCTCTCACTGATACATCAACCAAGATGATTTTAA 4125
 453 Thr.....LysLysIleGlyLysArgLeuAsnIleGlnLe 464
 : : : : :
 4126 ACTTTTCTTCATCACCCTCAAGCCGAGATATCTTTGAGGTGAACT 4175
 464 uLysLysGlyThrGluGlyLeuGlyPheSerIleThr.....Sera 478
 : : : : :
 4176 GGTAAAAATGATATACAGCTTGGGATAGTGCACGGAGGTGTGAATA 4225
 478 rGAspValIThrIleGlyLysSerAlaProIleTrpValLysAsnIleLeu 494
 : : : : :
 4226 CGAGTGCACACATGCTGGGC.....ATTATGTCAAAAGCTGTAA 4266
 495 ProArgGlyAlaIleAlaIleGlnAspGlyArgLeuLysAlaGlyAspArgLe 511
 : : : : :
 4267 CCCAGGAGCAGCAGCAGATCTGATGTGAAATTCACAAAGGTGATCCGT 4316
 511 uIleGluValAsnGlyValAspLeuValGlyLysSerGlnGluGlnVal 528
 : : : : :
 4317 CCTAGCTCAATGAGTGTAGTGAAGAGCACCATTACAGCAAGCTG 4366
 528 aSerLeuLeuArgSerThrLysMetGluGlyThrValSerLeuLeuVal 544
 : : : : :
 4367 TGAAGAACCTGAGAAATCA.....GACAGGTGTCTCATCTGTGA 4407
 545 PheArgGlnGluAspAlaPheHisProArgLys..... 555
 4408 TTAGAAAAGGAGCAATCTCCACATCTMAAGAACATGTCCCGGTAAACCC 4457
 556LeuAsnAlaGluProSerGlnMetGlnIlePro..... 566
 4458 ACAGTGTACCTTTCAGATCAGAAATCCCAAGGTCAAGGCCCAAGAAAAG 4507
 567LysGluThrLysAlaGluAspGluAspIleValLeuThrProAsp 581
 4508 TGAAGAAACAACTCAGTCAAGAAAGACTACAGCTTGTG..... 4545
 582 GlyThrArgGluPheLeuThrPheGluValProLeu...AsnAspSerG 597
 : : : : :
 4546ACTGAAGAAATATACATTTAGGTAATAATTAATAATACTGC 4589
 597 ySerAlaGlyLeuGlyValSerValLysGlyAsnArgSerLysGluAsnH 614
 : : : : :
 4590 AGCTTAGGATTCAGTTTCTCGAGAAAGATATCTTAACCGAGACAAA 4639
 614 IsAlaAspLeuGlyIlePheValLysSerIleIleAsnGlyAlaAla 630
 4640 TTAAATGCCAGCATAGTAAAGGTTAAAAAGCTCTTCTCGACAGCCAGCA 4689
 631 SerLysAspGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsnG 647
 : : : : :

```

4690 GCAGAAAGTGGAAAAATTGATGTAGAGATGTTATCTTGAAGTGAATGG 4739
647 YGUSerLeuLeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArg 664
      ||||| ||| :|||||:|||||:|||||:|||||
4740 AGCCTTTTGAAGACTATCTCAGCAGAGATCATATCGCTCTCAGG 4789
664 rGSeMetSerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleVal 660
      :||| :|||:|||||:|||||:|||||
4790 GAACCTCTCCAGAA.....GTATCTTGCTTC 4818
681 AlaArgArgIleSerLysCysAsnGluLeuLysSerProGlySerProP 697
      :||| :|||:|||||:|||||:|||||
4819 TCGAGA.....CCTCCACC 4832
697 OGlyProGluLeuProIleGluThrAlaLeuAspAspArgGluArgI 714
      ||||| :|||||:|||||:|||||:|||||
4833 TGGTGTGCTACCGCAAAATGATCTCGGCT..... 4863
714 leSerIleSerLeuTySerGlyIleGluGlyLeuAspGlySerProSer 730
      ||| ||| :||| :|||:|||||:|||||
4864 ..TTGACCCCACTTCAGTCTCCAGCACAAGTACTCCAAACACAGTAAA 4911
731 ArgAsnAlaIleLeuSerArgIle..... 738
4912 GACTCTTTCAGCATCATGTGTGGACAAAGCAGCAGCTCAGATGAAA 4961
739 ...MetGlyLysSerGlyLysTyGlnLeu...SerProThrValAsnM 753
      |||||:|||||:|||||:|||||:|||||
4962 TGAATGTCCAGCAAAAGCAAAACAGTCAGTCCCACTCCAGAAAG 5011
753 eGProGlnAspAspThrValIleIleGlnAspAspArgLeuProValLeu 769
      :|||||:|||||:|||||:|||||:|||||
5012 ACAGTTACAGTGCACAGTGGAGTGGAGAAAGATGACTTACAGACAGCT 5061
770 ProProHleLeuSerAspGlnSer...SerSerSerSerHisAspAsp 785
      ||| :|||||:|||||:|||||:|||||
5062 CCAGCAATATCATATTCGACCTGAGTGCAGCTTGCATCAGCTCT 5111
785 IglIleValThrAlaAspAlaGlyThrThrAlaLysAlaIleSer. 801
      :||| :|||||:|||||:|||||:|||||
5112 AAGCAACAGTGTATCAGCAGCAGCAGATCATGTAAGCAACCAAGCTC 5161
801 ..... 801
5162 AAGAGATACCATTTGTACCATGTTTACTATCAGAAATTTCCCAAT 5211
802 .....AspSerAlaAspCysSerLeuSerProAspValAs 813
5212 AAACGAGATTGTAGGACAGTAAATCTCCCTCTACCCAGCGATATGGC 5261
813 pProValIleuAlaPheGlnArgGluGlyPheGlyArgGlnSerMetSer 829
      ||| :|||||:|||||:|||||:|||||
5262 TCCTGGCAGAGTATCAACCCCAATCAGAAATCTGCTCTAGTTCCG 5310

```

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-410-804-2

seq_documentation_block:

```

: Sequence 2, Application US/08410804
: Patent No. 5632994
: GENERAL INFORMATION:
: APPLICANT: Reed, John C.
: APPLICANT: Sato, Takaki
: TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cathryn Campbell
: STREET: 4370 La Jolla Village Drive, Ste 700
: CITY: San Diego
: STATE: California
: COUNTRY: United States
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/410,804
: FILING DATE: 27-MAR-1995
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/259,514
: FILING DATE: 14-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 1389
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1830 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-410-804-2

```

```

alignment_scores:
  Quality: 275.50      Length: 533
  Ratio: 1.036         Gaps: 19
  Percent Similarity: 49.906   Percent Identity: 24.765

```

alignment_block:

US-09-757-781-2 x US-08-410-804-2 ..

Align seq 1/1 to: US-08-410-804-2 from: 1 to: 1830

```

360 AlAlaAsnLysGlnGlyThrGlnLeuSerGlnSerGlnLysAsn.. 375
      |||:|||||:|||||:|||||:|||||
34 GCCACCGAAGAGACTTTCAGTGAATGTAACCAAGCAAAAGTAAATA 83
376 .....AsnTyTyTySerSerArgPheSerProAspS 386
      :||| :|||||:|||||:|||||:|||||
84 GCCAGGATTTGATGTATCTGATTAATCTCAGACCGTGAGATTCAGAC 132
386 eGlnTyIleAspAsnArgSerValAsnSerIleGlyLeuHisThrVal 402
      :||| :|||||:|||||:|||||:|||||
133 .....ATGATGAAGCCACTTACTCCAGACTCAGATCAT..... 168
403 GlnArgAlaProArgLeuAsnHisProProGlnGlnIleAspSerHis 419
      :||| :|||||:|||||:|||||:|||||
169 .....CAACACCAAAACAG...GAATCTTCCTC 194
419 rArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaPro 436
      :||| :|||||:|||||:|||||:|||||
195 TTCAGTGAATATACATCCAAAGATGAATTTTAAACTTTTCTTCACAC 244
436 laSerAlaProGlnAsnValPheSerThrThrValSerSerGlyTyAsn 452
      ||| :|||||:|||||:|||||:|||||
245 CTCCTAAGCCTGGAGATATCTTT..... 267
453 ThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysGlyThrG 469
      :||| :|||||:|||||:|||||:|||||
268 .....GAGCTGAACCTGCCTAAATAATGATAA 293
469 uGlyLeuGlyPheSerIleThr.....Sera 478
      :||| :|||||:|||||:|||||:|||||
294 CAGCTTGGGATAGTGTCAAGGACTGTGTGACAGGAGAGTGTGATA 343
478 rGaSPValThrIleGlyLysSerAlaProIleTyValLysAsnIleu 494
      :||| :|||||:|||||:|||||:|||||
344 CGAGTGTACAGATGCTGCC.....ATTATGTGAAAGCTGTATT 384

```

```

495 ProArgGlyAlaAlaIleGlnAspGlyArgLeuLysAlaGlyAspArgLeu 511
385 CCCCAGGAGACACAGATCTGATGTAGATTAATTCACAAAGGATCGCGT 434
511 uileGlyValAsnGlyValAspLeuValGlyLysSerGlnGluVal 528
435 CCGAGCTGCATGAGTTAGTCTAGAACAGCCACCAATAGCAAGCTG 484
528 aSerLeuLeuArgSerThrLysMetGluGlyThrValSerLeuLeuVal 544
485 TGGAAACACTGAGAAATACA.....GGACAGTGTTCATCTGTTA 525
545 PheArgGlnGluAspAlaPheHisProArgGlu..... 555
526 TTGAGAAAGGACAAATCTCCAAATCTAAGAACATGTCCCGTAACCC 575
576 ACAGTGTACCTTTCAGATCAGAAATGCCCAAGGCTCAAGGCCAGAAAAA 625
567 ....LyGluThrLysAlaGluAspGluAspIleValLeuThrProAsp 581
626 TGAAGAAACAACTCAGCTCAAGACTACAGCTTGTG..... 663
582 GlyThrArgGluPheLeuThrPheGluValProLeu...AsnAspSerG 597
664 .....ACTGAAGAAATACATTGAGGTAAATTAATTAATAATGACTC 707
597 ySerAlaGlyLeuGlyValSerValLysGlyAsnArgSerLysGluAsn 614
708 AGGCTTAGGATTCTCTCCAGAAAGATTAATCTTATACCGAGAGAA 757
614 IsAlaAspLeuGlyIlePheValLysSerIleLeuAsnGlyAlaAla 630
758 TTAATGCCAGCATAGTAAGGTTAAAGCTCTTCCTGACAGCAGCAGCA 807
631 SerLysAspGlyArgLeuArgValAsnAspGlnLeuIleAlaValAsn 647
808 GCAGAAAGTGAAATGATGATGAGGATGATATCTTGAAGTGAATG 857
647 yGluSerLeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArg 664
858 AGCCTTTTGAAGCACTATCTCAGCAGCAAGCATATCTGCTCCAGCG 907
664 ySerMetSerThrGluGlyAsnLysArgGlyMetIleGlnLeuLeuVal 680
908 GAACCTGCTCCAGAA.....GATTCCTGCTTC 936
681 AlaArgArgIleSerLysCysAsnGluLeuLysSerProGlySerPro 967
937 TGCAGA..... 1000
697 cGlyProGluLeuProIleGluThrAlaLeuAspAspArgIuArgArg 714
951 TGGTGTGCTACCGGAATGTGATACCTGCTT..... 981
714 leSerHisSerLeuLysSerGlyIleGluGlyLeuAspGluSerProSer 730
982 ..TTGACCCCACTTCAGTCTCCAGCACAAGTCTCCAAACAGCAGTAA 1029
731 ArgAsnAlaAlaLeuSerArgIle..... 738
1030 GACCTCTCTCAGCCATCATGTGTGAGCAAAAGCAGCAGTCAATGAAA 1079
739 ...MetGlyGluSerGlyLysTyrGlnLeu...SerProThrValAsn 753
1080 TGAATGTGACACAAACCAAAAGAGTGCAGATCCCAATCCAGAGAG 1129
753 etProGlnAspAspThrValIleIleGlnAspAspArgLeuProValLeu 769
1130 ACAGTTACAGTACAGCAGCTGGAGTGAGAGAAAGTGAAGTGAAGCT 1179
770 ProProHisLeuSerAspGlnSer...SerSerSerHisAspAspVa 785

```

```

1180 CCAAGCAACATCAATCAATTCGACCTGAGTTCAGCTTTCGATCAGACT 1229
785 lGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIleSer 801
1230 AAGCAACATGTATCAAGCAGCAGCAGATCATGTAAGCAGCAGAGATC 1279
801 ..... 801
1280 AAGAATACCATTTTACCATGTTTACTATCTCAGAAAAATCCCAAT 1329
802 .....AspSerAlaAspCysSerLeuSerProAspValAs 813
1330 AAACCAAGTGTGAGCAGCAGATTCCTCCCTACCAACCGATATGCG 1379
813 pProValLeuAlaPheGlnArgGluGlyPheGlyLysArgIleSerMetSer 829
1380 TCCTGGGCGAGCTTATCAACCCCAATCAGAAATCTGCTCTGTAGTTCG 1428
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-259-514-2

```

```

seq_documentation_block:
; Sequence 2, Application US/08259514
; Patent No. 5747245
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: FAS ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cathryn Campbell
; STREET: 4370 La Jolla Village Drive, Ste 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/259,514
; FILING DATE: 14-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9954
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-259-514-2

```

```

alignment_scores:
    Quality: 275.50      Length: 533
    Ratio: 1.036        Gaps: 19
    Percent Similarity: 49.906    Percent Identity: 24.765

```

alignment_block:

US-09-757-781-2 x US-08-259-514-2 ..

Align seg 1/1 to: US-08-259-514-2 from: 1 to: 1830

360 AlaAlaAsnLysGluGlnTyrGluGlnLeuSerGlnSerGlnLysAsn.. 375


```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,311
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/410,804
FILING DATE: 27-MAR-1995
APPLICATION NUMBER: US 08/259,514
FILING DATE: 14-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1389
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-858-311-2

alignment_scores:
Quality: 275.50 Length: 533
Ratio: 1.036 Gaps: 19
Percent Similarity: 49.906 Percent Identity: 24.765

alignment_block:
US-09-757-781-2 x US-08-858-311-2 ..

Align seg 1/1 to: US-08-858-311-2 from: 1 to: 1830
360 AlaAlaAsnLysGluGlnTyrGluGlnLeuSerGlnSerGlnLysAsn.. 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 GCCACCGGAAAGAGACTTTCACGTAGTAAACCAAGCAAACTATAAAA 83
376 ..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::
84 GCCAGCATTTTCGATGTAAGTACGTACGACCGGTGAGATTCAGAC.. 132
386 ergLntYrllleAspAsnArgSerValAsnSerAlaGlyLeuHlSerHlVal 402
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 ..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::
403 GlnATGAlaProArgLeuAsnHlSerProGluGlnLleAspSerHlSe 419
169 ..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::
419 fArgLeuProHlSerAlaHlSerProSerGlyLysProProSerAlaProA 436
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 TTGAGGAATACATCCACACAGATGAATTTTAAACTTTTCTTCATAC 244
436 lAserAlaProGlnAsnValPheSerHlThrValSerSerGlyTyrAsn 452
245 CTCCTAACGCTCGAGATATCTTT ..:::..:::..:::..:::..
453 ThrLysLysLllegLysArgLeuAsnLllegLlnLeuLysGlyThrGl 469
268 ..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::
469 uGlyLeuGlyPheSerLleThr ..:::..:::..:::..:::..:::
294 CAGCTTGGAATAGTGTACAGTACGTTGTCACAAAGAGGTGTGTAA 343
478 rGaAspValThlLllegLysSerAlaProLleTyrValLysAsnLleLeu 494
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

344 CGAGTGTACAGATGTTGCC.....ATTATGTGAAGGTGTATT 384
495 ProArgGlyAlaAlaLllegLlnAspGlyArgLeuLysAlaGlyAspArgLe 511
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
385 CCCAGGAGCAGCAGAGTCTGATGTTAGTAATTCACAAAGGTATCCGCT 434
511 uLlegLlnValAsnGlyValAspLeuValGlyLysSerGlnGlnLlnVal 528
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 CTTGCTGTCAATGAGGTGTGTAGAGAGAGCCACCAATAGCAGACG 484
528 alSerLeuLeuArgSerHlThrMetGlnGlyThrValSerLeuVal 544
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 TGGAAACACTGACAAATAC.....GGACAGGTGTCTATCTGTTA 525
545 PheArgGlnGlnAspAlaPheHlSerProArgLln..... 555
526 TTGAAAAGGACAAATCTCAACATCTAAAGAACATGTCCGGTACCC 575
556 ..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::
576 ACAGTGTACCTTTCAGATCAGAAATGCCAAGGTCAAGGCCAGAAAAAG 625
567 ....LysGlnThrLysAlaGlnAspGlnAspLleValLleuThrProAsp 581
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
626 TGAAGAAAACACTCAGGTCAAGCTCAAGCTTGTTC..... 663
582 GlyThrArgGluPheLeuThrPheGlnValProLeu..AsnAspSerG 597
664 ..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::
597 ySerAlaGlyLeuGlyValSerValLysGlnAsnArgSerGlyLlnAsn 614
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 AGCTTACGATTCAGTTTCTCGAGAGATTAATCTTATCCGGACAA 757
614 lAlaAspLeuGlyLlePheValLysSerLleLlnGlnGlyAlaAla 630
758 TTAATGCCAGCATGTAAAGGTTAAAGCTTTCTTGTGACAGCAGCA 807
631 SerLysAspGlyArgLeuArgValAsnAspGlnLleLlnAlaValAsnG 647
808 GCAGAAAGTGGAAAATGTAGTGAAGAGATGTATCTGAAAGTGAATGG 857
647 yGlnSerLeuLeuGlyLysThrAsnGlnAspAlaMetGlnHlLeuArg 664
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
858 AGCCTTTTGAAGGACATATCTCAGCAGAAATCATCTCTCTCAGGG 907
664 rGserMetSerHlThrGlnGlyAsnLysArgGlyMetLllegLlnLleVal 680
908 GAACCTCTCCAGAA.....GTAATCTTCTCTCTC 936
681 AlaArgArgLleSerLysCysAsnGlnLeuLysSerProGlySerProP 697
937 TGCAG.....CTCCACC 950
697 oGlyProGlnLeuProLllegLlnThrAlaLeuAspAspArgGlnArgArg 714
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 TGTGTGCTACCGGAATGTGATCTGCGCTT..... 981
714 lSerHlSerLeuLysSerGlyLllegLlnGlyLeuAspGlnSerProSer 730
982 ..TTGACCCCACTTCAGTCTCCAGCACAACTACTCCAAACACAGTAA 1029
731 ArgAsnAlaAlaLeuSerArgLle..... 738
1030 GACTCTTCTCAGCCATCATGTGTGGAGCAAAAGCAGCAGTCAATGA 1079
739 ....MetGlyLysSerGlyLysTyrGlnLeu..SerProThrValAsn 753
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1080 TGAATGTCTCAGACAAAAGCAAGTGCAGATCCCATCCAGACAG 1129
753 etProGlnAspAspHlThrValLllegLlnAspAspArgLeuProValLeu 769
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1130 ACAATTACGTGACAGCACTGGAGGTGAGAGATGACTTGTAGTACAGCT 1179

```

```

770 ProProHISLeuSerAspGlnSer...SerSerSerSerHISAspAspVa 785
1180 CGAGCAAAATATCAATTCGACCTGAGCTTTCATCAGACACTCT 1229
785 IGLYPheValThrAlaAspAlaGLyThrTrpAlaLysAlaAlaIleSer. 801
1230 AAGCAACATGTATCACAGCAGATCATCATGAAACCAACCAAGAGTC 1279
801 ..... 801
1280 AAGAATATCATTTGTACCATGTTTACTATTCCTCAGAAATTCCTCAT 1329
802 .....AspSerAlaAspCysSerLeuSerProAspValAs 813
1330 AACCAGAGCTTTCAGCAGACAGTATCTCCCTCCCTACACACCGATATGCG 1379
813 ProValLeuAlaPheGlnArgGluGlyPheGlyArgGlnSerMetSer 829
1380 TCCTGGCGCAGAGTATCAACCCCAATCAGAAATCTGCTTCTCTAGTTCG 1428

seq_name: /cgn2_6/ptodata/2/lna/6a_COMB.seq.us-08-931-999-4

seq_documentation_block:
: Sequence 4, Application US/08931999
: Patent No. 6043219
: GENERAL INFORMATION:
: APPLICANT: Iandolo, John J.
: APPLICANT: Crupper, Scott S.
: TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hovey, Williams, Timmons & Collins
: STREET: 2405 Grand Boulevard, Suite 400
: CITY: Kansas City
: STATE: Missouri
: COUNTRY: U.S.A.
: ZIP: 64108
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/931,999
: FILING DATE:
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/710,561
: FILING DATE: 19-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Collins, John M.
: REGISTRATION NUMBER: 26,262
: REFERENCE/DOCKET NUMBER: 25043-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 816/474-9050
: TELEFAX: 816/474-9057
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6755 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Staphylococcus aureus
: STRAIN: UT0007
: US-08-931-999-4

```

```

Quality: 244.50      Length: 1306
Ratio: 0.414      Gaps: 56
Percent Similarity: 45.253      Percent Identity: 19.219

alignment_block:
US-09-757-781-2 x US-08-931-999-4 ..

Align seg 1/1 to: US-08-931-999-4 from: 1 to: 6755

154 SerValSerAspSerAsnPheSerSerGluGluProSerArgLysAsnPr 170
111 ..... 111
610 AGCAGAGCAAAACAAAGAAAGGAGGAGCACCACGAAACGAAAAAAG 659
170 othrArGTPSerThrThralaGlyPheLeuLysGlnAsnThrala.... 185
660 AACAAAA...AAGAAAGAACCCGAAAGCCGAAAGAAAGAAAGAAACAA 706
186 .....GlySerProLysThrCysAspArg 193
707 AAGGCAAAACGACCCACACAGAAAAACACACCAAAAGAGACAGA 756
194 LysLysAspGluAsnTyArgSerLeuProArgAspThrSerAsnTrpSe 210
757 AAAAAAAGCAAAAGAAAAAAGAAAA...CGAAGCAACAGAAACCCAG 800
210 rAsn.....GlnPheGlnArgAspAsnAlaArgSerSerLeuSerAla 225
801 CAACAAAAAACACAGCAAGAAAGACGAAAGAAACAAACAAACACACCCCA 850
225 erHisProMetValGlyLysTrpLeuGluLysGlnGluGlnAspGluAsp 241
851 AAACAAAGAAAGCAACCAAA...AAAAAGCAAAACCAACCAAAAG 894
242 GlyThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspTh 258
895 GGCACGCAAGAGAGAGAAAAAGC..... 918
258 rGlyLeuGluHisIleProAsnPheSerLeuAspAspMetValLysLeuV 275
919 .GGAAAGCAAAAGAAACCA..... 936
275 alGluValProAsnAspGlyGlyProLeuGlyIleHisValAlaProPhe 291
936 ..... 936
292 SerAlaArgGlyGlyArgThrLeuGlyLeuValLysArgGluL 308
937 .....AGGAAAGGACCAAAAGAAAGAAAAAAGAAAAAAGCGAGAAA 980
308 ysgLysGlyLysAlaGlnHisGluAsnLeuPheArgGluAsnAspCysIle 324
981 CAGGAAAAAAGAAAGAAACAAACAGAGAAAGCAACAAACGAAACACGG 1030
325 ValArgIle.....AsnAspGlyAspLeuArgAsnAlaGlyPhe 337
1031 ACCGAAACCAAAAAAAGAAAGCAAGCCG....AGGAAAGAAAGA 1074
337 eGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThrProIleLeu 354
1075 CAAGAAAAAAGCAAGAGACAGACAGCGAGAGAGAAAC..... 1114
354 rPheHisValValProAlaAla..... 361
1115 .....CACAAAGAAACAAAGACAGAGAAACGAAAGCAAGCAACCA 1159
362 .....AsnLysGluGlnTyArgGluGlnLeuSerGlnSerGly 374
1160 CCCCAAGAAAGAAAGAAAGAAAGCAAGAGAAAGCAAGCAAGAGAA 1209
374 sAsnAsnTyArgSerArg..... 381
1210 AAAAAAGCGGCAAAAGCAAGACCAAGAGAGCAAGCAAGAAAGAAAGAA 1259

```

alignment_scores:

Thu Jul 25 08:38:38 2002

```
382 ..... pheSerProAspSerGlnTyrIleAsp 390
1260 GAGGACGAAAGAAAACAGAGACACCAAGAACGAGAAAGAA 1309
391 AsnArgSerValAsnSerAlaGlyLeuHisThrValGlnArg .....Al 405
.....
1310 AGAAGACCGGAAACCGCAAAACGAGACCGCCAAAGGAGAGAG 1359
405 AspArgLeu.....AsnHisProProGlnGlnIleAspSer 417
.....
1360 ACCCAGAGAGGACGAAAGGCGAAGCGCCCGCAACGAAACCA 1409
418 .....HisSerArgLeuPro 422
.....
1410 AAAAAAACCAAGAAAGAGCGAGGAAAGCAAAAGAGAGAA 1459
423 HisSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaPr 439
.....
1460 CAACGACACAGAAACACCGGAGAGCGCGAAGCGCAAGAGCC.. 1507
439 GlnAsnValPheSerThrValSerSerGlyTyrAsnThrLysIle 456
.....
1508 .....AACGGGAAAGACACGAAACCAAGGAA 1538
456 IAGLYLysArgLeuAsnIleGlnLeuLysGlyThrGlnGlyLeuGly 472
.....
1539 ACAGCAAAAGCGAAAGAGAAAGAAAGAAAG 1570
473 PheSerIleThrSerArgAspValThrIleGlySerAlaProIleTyr 489
.....
1571 .....AAGAAAGCGCGAGCAAGACAGGAGAGCGGAGCGCAAA 1608
489 rValLysAsnIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeu 506
.....
1609 GAGAAAGAAAC.....CAAGAAAGACCGACGA 1634
506 ysaIaGlyAspArgLeuIleGlnValAsnGlyValAspLeuValGlyLys 522
.....
1635 AATCAAAAGAAAG.....GAAAG 1654
523 SerGlnGlnGlnValValSerLeuLeuArgSerThrLysMetGlnGlyTh 539
.....
1655 AAGCAAAAG.....AAAAAGAAAGAACACCGGAG 1686
539 rValSerLeuLeuValPheArgGlnGlnAspAlaPheHisProArgGlu 556
.....
1687 CCAGGAAAGCCGAGAGCAAGCAAAAGAAAGCAAAAGAAAGAA 1736
556 euAsnAlaGlnProSerGlnMetGlnIleProLysGlnThrLysAlaGlu 572
.....
1737 AATAGAGAACAAAGAAACAAAGCAAGCAAGCAAAAGAAAGCGGAA 1786
573 AspGlnAspIleValLeuThrProAspGlyThrArgGlnPheLeuThrPh 589
.....
1787 GACAAA..... 1792
589 eGlnValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValL 606
.....
1793 .....A 1793
606 ysaGlyAsnArgSerLysGlnAsnHisAlaAspLeuGlyIlePheValLys 622
.....
1794 AATACACGCGGCAAGCGAGACACGCAACGAAAGGA..... 1832
622 sSerIleLeuGlnGlyAlaAlaSerLysAspGlyArgLeuArgValA 639
.....
1833 .....AAAGCGGAAACAAAGAGAAAGAAAGAAAGCAAGAA 1872
639 snAspGlnLeuIleAlaValAsnGlyGlnSerLeuLeuGlyLysThrAsn 655
.....
1873 ACCAGAGCGGACCGCAGGAGGAGAGAGAAAGAAAGCAAGAGCGAA 1922
656 GlnAspAlaMetGlnThrLeuArgArgSerMetSerThrGlnGlyAsnLys 672
```

us-09-757-781-2.rni

```
1923 AAG...AAACACGAAAGAAAGAAAGAAAGCGAGCGCGCAAGAAAGAAACA 1969
672 sarGlyMetIleGlnLeuIleValAlaArgArgIleSerLysCysAsnG 689
.....
1970 AAGAGGA.....GAGG 1980
689 IuLeuLysSerProGlySerProProGlyProGlnLeuProIleGlnThr 705
.....
1981 AGAATACACGCCAGGA..... 1997
706 AlaLeuAspAspArgGluArgArgIleSerHisSerLeuTyrSerGlyI 722
.....
1998 .....AAGACAGGAAAGACCGGA..... 2015
722 eGlnGlyLeuAspGlnSerProSerArgAsnAlaAlaLeuSerArgIle 739
.....
2016 .....GAACAGAGAAACAGAGGAAAGAGCGCCCAACAGAGGA 2055
739 etGlyGlnSerGlyLysTyrGlnLeuSerProThrValAsnMetProGln 755
.....
2056 AAGGCAAGAAAGAAAGAAAGAAAGAGCGCCCAACAGAGGA 772
756 AspAspThrValIleIleGlnAspAspArgLeuProValLeuProOH 784
.....
2079 .....SerSerSerHisAsp 784
772 sLeuSerAspGlnSer.....SerSerSerHisAsp 784
.....
2090 CAAGCGCAACAGAGAGCGAGCGGAAAGAAAGCAAGAAAGCGACCGA 2139
784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
.....
2140 ACAGAAAGAAAGAAAGCAAGAGAGAGGAGAAACAGAAAGGGGGA 2189
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
.....
2190 GAACCAAGAGAGAGAAAGAAAGAAAGCGCAAGAGAGCG..... 2231
817 aPheGlnArgGlnGlyPheGlyArgGlnSerMetSerGlnLysArgThrL 834
.....
2232 .....AAGCGGAAAGCGGACAGCGAGAGCA 2259
834 ysaGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
.....
2260 AGAAGAAAGAGAGAGAGCAACAGAGAGAGAAAGAAAGCAAGCA 2309
851 LysSerMetAspLeuGlyIleAlaAsp..... 859
.....
2310 AGAAGAAAGAG.....GACCAAGAGAAAGCAAGCAAGCA 2347
860 .....GlnThrLysLeuAsnThrValAspAspGlnLysAlaGlyS 873
.....
2348 CGGACAAAGAAAGAAAGCAAGAGCGCAAGAGAAAGAAAGAGCA 2397
873 erProSerArgAspValGlyProSerLeuGlyLeuLys.....Lys 886
.....
2398 GGAAGCAAGCAACAGCAACCAAGCCACACAAAGAAAGAGAGCA 2447
887 SerSerSerLeuGlnSerLeuGlnThrAlaValAlaGlnValThrLeuAs 903
.....
2448 CCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2486
903 nGlyAspIleProPheHisArgProArgProArgIleIleArgGlyArg 920
.....
2487 .....AAGCGCAACAAAGAAAGAAAGAGAA 2514
920 LysCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrAspLysPro 936
.....
2515 GGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2549
937 AlaValAspAspAspAspGlnGlyMetGlnThrLeuGlnGlnAspThrG 953
.....
```



```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-045-632-1

```

```

alignment_scores:
    Quality: 233.50      Length: 1056
    Ratio: 0.509         Gaps: 43
    Percent Similarity: 43.466    Percent Identity: 19.697
alignment_block:
US-09-757-781-2 x US-09-045-632-1

```

Align seg 1/1 to: US-09-045-632-1 from: 1 to: 4527

```

215 ArgAspAsnAlaArgSerSerLeuSerAlaSerHisPrometValGlyLy 231
|||.....|.....|.....|.....|.....|.....|.....|.....|
685 CGAGGGGGAGCGCATGATGACAGAAATAAATGCCCGCTGCGATGATMAC 734
231 sTrpLeuGlyLys...GlnGluGlnAspGlyThrGluGlnAsp 247
.....|.....|.....|.....|.....|.....|.....|.....|
735 CTGTGTCTGCTCGAGGGCGCTGCTGACAGAGAGGCGACCATCAACCTG 784
247 snSerArgValGluProVal...GlyHisAlaSprThrGlyLeuGlnHis 262
.....|.....|.....|.....|.....|.....|.....|.....|
785 GAGACAGGTGCTCAGTGTGTGATGGAATTCGGCTTCGGGAACCAACCCAT 834
263 IleProAsnPheSerLeu.....AspAspMetValLysLe 274
.....|.....|.....|.....|.....|.....|.....|.....|
835 GCTAGAGCGCATGACATCTTAAACAGTGGAGCAAGAGCAACCGCTCT 884
274 ValGlu.....ValProAsnAspGlyGlyP 283
|||.....|.....|.....|.....|.....|.....|.....|.....|
885 GATGATATGATGATGTCCTCGTATGATGATTCGTGACACAGACATCCGGC 934
283 rOlouGlyIleHisValAlaProPheSerAlaArgGlyArgThrLeu 299
|||.....|.....|.....|.....|.....|.....|.....|.....|
935 CACTA.....CTAGTTGAAGTTGCCAAACCTCCGGGTGCACACCTT 975
300 Gly.....LeuLeuVal 303
976 GGGGTTCGACTAACTACCTCGGTGCTGTACCAACACAGTCACTGTCAT 1025
303 lLysArgLeuGlyLysGlyLysAlaGlnHisGlnAsnLeuPheArg 320
:.....|.....|.....|.....|.....|.....|.....|.....|
1026 AGACAAATCAAAATCTGCAAGCATTCGGGACAGTGGGGGGCGCTACATG 1075
320 lUAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnArg 336
.....|.....|.....|.....|.....|.....|.....|.....|
1076 TGGGAGACCAACATCTCTCCATCGACGCCAGCGATGAGTACTGTACC 1125
337 PheGluGlnAla.....LeuVal 340
:.....|.....|.....|.....|.....|.....|.....|.....|
1126 CTCGAGAACCAACCACTTCCTGGCCAAACACACTGACAGGCAAGCT 1175
341 .....GlnHisMetPheArgGlnAlaMetArgThrProIle 1175
1176 GGGAGATTCTCCACACCATGACAGACCCGCTGCGCTGAGAGGACCTGA 1223
353 lETrpPheHisValAlaProAlaAlaAsnGlyGlnIleArgGlnLeu 369
1224 .....CATGTCAAAAATTCAAGAGCGACAG..... 1250
370 SerGlnSerGlyLysAsnAsnTyrTyrSerArgPheSerProAspSe 386
1250 ..... 1250
386 rGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHisThrValG 403
1251 .....ACAGCTTCCTGGGATCTCGGGCCAGCAGCAACATGCT 1287

```

```

403 lNArgAla..ProArgLeuAsnHisProProGluGlnIleAspSerHisSe 419
|||||.....|.....|.....|.....|.....|.....|.....|.....|
1288 AGCGTTCATATACCAACCATACCATATACCCGACCCACGAC...CATGG 1334
419 rArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaPro. 435
|||.....|.....|.....|.....|.....|.....|.....|.....|
1335 CCGAGTACACAGCCCTGGGTTCCTCGAAAGCGCTTACTCCAAACAGCCCTC 1384
436 .....AlaSerAlaProGlnAsnValPheSerThrThrVal 447
1385 CGCGATGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1434
448 Ser.....SerGlyTyrAsnThrLysLy 455
|||.....|.....|.....|.....|.....|.....|.....|.....|
1435 AGTTCCTCGAATGGGACTTACTCTGACGACCTCTCTCTCTCTCTCTCTCT 1484
455 sIlleGlyArgLeuAsnIleGlnLeuLysLysGlyThrGluGlyLeuG 472
|||.....|.....|.....|.....|.....|.....|.....|.....|
1485 ACAGAGAACCATGATGAGAGAGAGACTGAAAAAGAGACTTCAAAGCT 1534
472 lPheSerIleThrSerArgAspValThrIleGly..... 484
1535 CACTGTCTTACCTCTAGCACTGTGGATGTGGCTGACAGGTGTTCAC 1584
484 ..... 484
1585 ACTGAACACACAGAGCTTGTCTGACGCTGATCTCTCTCTCTCTCTCTCT 1634
485 .....SerAlaProI 488
1635 AATCCACCTGAGGAGCGCTGTGTCACACAGACTCTCTCTCTCTCTCTCT 1684
488 lETrpValLysAsnIleLeuProArgGlyAlaIleGlnAspGlyArg 504
.....|.....|.....|.....|.....|.....|.....|.....|.....|
1685 CTGTGATTCCTATTTGAAGCTGACAGCGCCAGCAGAGAGATGTGGTGTG 1734
505 LeuLysAlaGlyAspArgLeuIleGluValAsnGlyValAspLeuValG 521
|||.....|.....|.....|.....|.....|.....|.....|.....|
1735 CTGCACATTTGGAGACAGACAGTCACTGATTAATGAATCCCAACAGAGA 1784
521 lYlSerGlnGluGluValAlaSerLeuLeuArgSerThrLysMetGlu 538
.....|.....|.....|.....|.....|.....|.....|.....|.....|
1785 CAGCACCTTTGAGAAAGCAATACAGTCTCTCTCTCTCTCTCTCTCTCTCT 1834
538 lYThrValSerLeu.....LeuValPheArg 546
:.....|.....|.....|.....|.....|.....|.....|.....|
1835 GCAAGTCACTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1884
547 GlnGluAspAlaPheHis.....ProArgGlnLeuAsnAlaGlu.. 559
.....|.....|.....|.....|.....|.....|.....|.....|.....|
1885 AGTATGTGAACATTTCAATGTAACCTGCTTAAGAAAGCAACAGCTGTGACT 1934
560 .....ProSerGlnMetGln..... 564
1935 TGGAAATTAACCATCAGTTCGCAATCCAGTAAGAAACACAGGGAGCCCTCG 1984
565 .....IleProLysGlnThrLysAla..... 571
1985 TCATTTCAATATCAAGAAAGGAGCTGTGGACACACAGACTGGAACCTCG 2034
571 ..... 571
2035 GAACCTTGAGATTAATTCCTGCGCATGATACATCCGGTTGGACAGCTG 2084
572 .....GluAsp.....GluAspIleVal. 577
2085 TTCCATGGAAGATGACATCCAGATCTCCAGAGTGTGAAGACTGTGGA 2134
578 .....LeuThrProAspGlyThrArgGlu..... 585
2135 AGCTCAAAATCCCAAAAGATGAAGATTACTCAGATGAGCAAGAGATTC 2184
586 ...PheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLe 601

```

```

2185 GGGCGCATTTATACAGGTGAGCTGAAGCCCTATGGGGGCC...CT 2221
2186 ugiValSerValIsgIySnaRgSerLySgluASnHlaAspLeug 618
2232 TGGCATCACCATTTCGGAAACG.....GAAGACCCCTTGAT...C 2269
618 lyllePheValIySerIleIleASnIyglYAlaIaSerLySAspGly 634
2270 CTAATTATCATCTGAGCCCTCAATAAGGGGAGATTAGCTGAAGACAGCA 2319
635 ArgLeuArgValaIaSnAspGlnLeuIleAlaValaIaSnIyIleuLe 651
2320 GGCATCCACATCGAGATCGAATCCATCCATCAATGACATGAGCTTCAA 2369
651 ugiLySThrASnGlnaSPAlaMetGluThrLeuArgArgSerMetSer 668
2370 GGGGAAAGCCCTGAGTGAAGCCATCCACTGCTACAGATGGGGGAGAGA 2419
668 hrgIuGlyASnLySArgGlyMetIleGlnLeuIleValaIaArgArgIle 684
2420 CT.....GTACCCCTGAAATTT 2436
685 SerLySCysASnGluLeuLySserProGlySerProGlyProGlyIle 701
2437 AAGAAGCAGACAGATGCTCAACCTGCTCAAGTCC.....AAGAAGCT 2480
701 uPfoIleGluThrAlaLeuASpASpArgIuArgArgIleSerHisSerL 718
2481 GCCCATCCCC.....AGCCACTCAA 2500
718 euThySerGlyIleGluGlyLeuASpGluSerPro...SerArgASnAla 733
2501 GTGACCTGAGATGGTGAAGAGAGACCCCTCCCAATACAAAGGCTGCG 2550
734 AlaLeuSerArgIle..... 738
2551 AAGCTCTGTACGTGACCCCTCCACAGTACCCAGCCTGAGAGTCTGT 2600
739 .....MetGlyIuSerGlyLySTyrgIuLeuSerProT 750
2601 GGACTCTCTGGAGAGGATCTGGAATAGATGCCAGGTATGGAGTCAAGCA 2650
750 hrValASnMetProGlnaSPASpThrValIleIleGluASpASpArgLeu 766
2651 CGACTTTTCAGACTTCAGGTACATTTCAACACCTGATGACTGAGAGAGT 2700
767 ProValLeuProProHisIleuSer.....AspGlnSerSerSerSer 781
2701 CCGAAGAAAGAGCCAGCCCTGTCCCGGTCGCCAAGCCTCGAAGCCAGAC 2750
781 rHISASpASpValaIyPheValaThrAlaASpAlaGlyThrTrpAlaIySA 798
2751 ATACCCAGATGTGGCTTGAGTAAATGAAGAC.....TGGAGCCGCT 2791
798 lAlaIleSerASpSerAlaASpCysSerLeuSerProASpValaSPPro 814
2792 CCACAGCCACTGGCTTCGCGGGGCTTCTGACAGTGCAGACGCTGAA... 2838
815 ValLeuAlaPheGlnArgGluGlyPhe..... 823
2839 .....CAAGAGAAACTTCTGGTCTCAACCATTTGAGAGACCT 2876
823 ..... 823
2877 GGAGACCTGGGGCAGTCGGGAATCTGAGAGAGCTCGAGCAACATCA 2926
823 ..... 823
2927 TGTGCGGAGCACTAGAGTTTGAATCATGAGGCTCAACGCTCGCAGT 2976
824 .....GlyArgGlnSerMetSerGluLySArgThrLySglnPheSerAS 838

```

```

2977 CAGCTGGGGCGCAGAGCCAGCTTCCAGAACGAGCAAT..... 3015
838 PaIaSerGlnLeuASpPheValIySThrArgLySerLySerMet.... 853
3016 ....TTCAGGGCCACATATACCAAGACACTTCGGACAGCAACTGCCCC 3061
854 ..AspleuGlyIleAlaSPGluThrLySLeuASnThrValaSPAsp... 868
3062 CGGATGTGGCGAGAAAGCTGTGACCTCGGAAATAATGACAAAGAAATC 3111
869 GlnLySAlaGlySerProSerArgASpValGlyProSerLeuGlyLeuLy 885
3112 AAGGAGATTATGTCCCAACTCCCGTGAGCTTCAAGAGTGACCTTATA 3161
885 slySerSerSerLeuGlnSerLeuGlnThrAlaValaIaGluValaThrL 902
3162 CAAGACTCTGCGATGAGAGACTTGGGTTGATGATGACATGAGTGGCTGC 3211
902 euASn.....GlyAsp 905
3212 TGGAGAAAGGGGTATGTCAAAATATCCGCCAGCTGGGCCAGGTGAT 3261
906 IleProPheHisArgProArgProArgIleIleArgGlyArgLyCysAS 922
3262 CTTCGAGGCTTGAAGCCTTATGACAGGCTCTTTACAG.....GTTAA 3302
922 nGluSerPheArgAlaAlaIleASpLySerTyASpLySProAlaVala 939
3303 CCAGCTCCGAACGAGAGACTCGACTGCTGCTGTGTGCTGCCCTCATNG 3352
939 SPASpASpASpGluGlyMetGluThrLeuGlnGluASpThrGluGluSer 955
3353 CCGAATCCGGTAACAGACTGAGCTGTTATTACAGAAACCCGCTGACC 3402
956 SerArgSerGlyArgGluSerValSerThrAlaSerAsp..... 968
3403 TCCCAAGATGATAGAACACCGGCTCTGCCAGTGTAGTGCAGACGAGCA 3452
969 .....GlnProSerHisSerLeuGluArgGlnMetA 979
3453 GAACAGGCTTCTTCAGACAGCCAGCCAGCAC.....G 3484
979 snGlyASnGlnGluLySgLyASpLySThrAsp..... 989
3485 GTGTAACCTGAGACAGACAGAACCCACTAATACATTAATGATTAATTT 3534
990 .....ArgLyLySASpLySThrGlyLySgIuLySlySlySA 1002
3535 TATAAGCAGGAGCAGAAAGCATATCTAATGATGTCTTAAACAAAGAGA 3584
1002 SPARGASpLySgIu 1006
3585 ACAGACAGAAACAG 3598

seq_name: /cgn2_6/ptodata/2/lna/5A.COMB.seq:us-08-056-200-93
seq_documentation_block:
; Sequence 93, Application us/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; NUMBER OF INVENTIONS: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA

```

```

? COUNTRY: U.S.A.
? ZIP: 92660
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patientin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/056, 200
? FILING DATE: 30-Apr-1993
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Fedrick, Michael F.
? REGISTRATION NUMBER: 36,799
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (714) 760-0404
? TELEFAX: (714) 760-9502
? INFORMATION FOR SEQ ID NO: 93:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 9551 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1507..1644
? NAME/KEY: Intron
? LOCATION: 1645..2511
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 2512..8070
US-08-056-200-93

alignment_scores:
      Quality: 216.00      Length: 1219
      Ratio: 0.391      Gaps: 50
      Percent Similarity: 45.283      Percent Identity: 19.688

alignment_block:
US-09-757-781-2 x US-08-056-200-93 ..

Align seg 1/1 to: US-08-056-200-93 from: 1 to: 9551
304 LysArgLeuGluLysGlyLysAlaGluHisGluAsnLeuPheArgG1 320
      ::::::::::::::::::::||| |||||::::::::::
4426 GAGAGGCTCGAGCGCGCTGAAGCGCGCATGAGAAAGAGAGCGCGA 4475
320 uAsnAspCysIleValArgIleAsnAspLysPheLeuArgAsnArgP 337
      :::::||||| :::::::
4476 GCAAG.....GAGCTACGTGAGGAGAGG 4498
337 heGluGlnAlaGlnHisMet.PheArgGlnAlaMetArgThrProIle1 353
      |||||:::::::::::|||| ||| ||:::
4499 AGGAAAGAGCGCGGAGCGAGATTAAAGCCCGCATCCCGAAGTGGCACTG 4548
353 eTrpPheHisValValProAlaAlaAsnLys...GluGlnTyrGluGlnL 369
4549 CAGCTAAGAAAGAGAGCGCGAGCGAGCAAGCAAAAGATCTTACTCGAGCG 4598
369 euSerGlnSerGluLysAsnAsnLysTyrTyrSerSerArgPheSerProAsp 385
4599 CCGCGAACAGAGAGGAGGAGCGCGCGCAAGAGAGAGGAGAAAGAGAGC 4648
386 SerGlnTyrIle.....As 390
4649 GGCGGAGAGTGAAGTCATGCGAGAGAGAGAGAGAGCGGCTCACCGGAG 4698

```

```

390 pAsnArgSerValAsnSerAla.GlyLeuHis...ThrValGlnArgAla 405
      ::::::::::::::::::::||| |||||::::::::::
4699 CACACAGAGAGAGAGAGAGCGCGGAGCTTCACATGAGCGAGCGGAG 4748
406 ProArgLeuAsnHisProProGluGlnIleAspSerHisSerArgLeuP 422
      ::::::::::::::::::::||| |||||::::::::::
4749 GGAAGAGAGAGAGAGCGCGCTCAGAGCGCTGTGGCAGAGCCCATTCG 4798
422 ohSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaP 439
4799 GCGAGCGAGAGAGAGAGCGAGCGCGGAGAGCGCGAGCGCGAGAGCG 4844
439 roGlnAsnValPheSerThrThrValSerSerGlyTyrAsnThrLys 455
4845 .....GAGACAGCGTTTCTCCCGA..... 4865
456 IleGlyLysArgLeuAsnIleGlnLeuLysIleGlnGlyThrGluLeuG1 472
4866 .....GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4885
472 yPheSerIleThrSerArgAsp.....ValThrIleGlyLys 485
4886 GCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4934
485 eAlaProIleThrValLysAsnIleLeuProArgGlyAlaAlaIleGln 501
4935 .....GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4952
502 AspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyValAs 518
4953 GCGGAGCGGTGCCCAACAGCTCCAGCA.....GAGAGAGAGAG 4990
518 pLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerThrL 535
4991 GCCTCCAGAGAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5040
535 yMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAlaPhe 551
5041 AATTC.....GAGGTGCGCACT 5057
552 HisProArgGluLeuAsnAlaGluProSerGlnMetClnIleProLysG1 568
5058 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5107
568 uThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrArg 585
5108 TACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5150
585 IuPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLeu 601
5151 .....GAGAGA 5156
602 GlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuG1 618
5157 GAGGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5203
618 yIlePheValLysSerIleLeuAsnGlyAlaAlaSerLysAspGlyA 635
5204 A.....GAGCAATATCCGCGAGAGAGAGAGAGAGAGAGAGAGAG 5244
635 rGluLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu... 650
5245 CAGCAGCTGCTGAG.....AGAGAGAGAGAGAGAGAGAGAGAGAGAG 5282
651 .....LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArg 665
5283 CAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 682
665 rMetSerThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAla 698
5333 AAGAGAGAGCTGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5360
682 rGArgTLeSerLysCysAsnGluLeuLysSerProGlySerProGly 698

```



```

1195 sServalServalGluValGlnMetGlnArgGln..... 1206
      : : : : : ||| ||| : : : : : |||
7029 ATTCATGAGAGAGAGAACAGCAGCTGCGCCGACGAGAGGCCACMACAGC 7078
      : : : : : ||| ||| : : : : : |||
1207 ..ArgGlnGlnGlnArgGlnSerSerGlnGlnGln..... 1218
      : : : : : ||| ||| : : : : : |||
7079 TCGCGCAGAGAGAGAGAACAAATTCGCCGAGAGAACAGCTGCTCCAGAA 7128
      : : : : : ||| ||| : : : : : |||
1219 ArgGlnTyrSerSerLeuProArgGlnSerArg..... 1229
      : : : : : ||| ||| : : : : : |||
7129 AGGGAAGAACAGCAGCTCCACCGCCAGAGAGCTGACAGAAATTCCTCA 7178
      : : : : : ||| ||| : : : : : |||
1229 ..... 1229
7179 GGAGAGAACCGAGCTGCCGCCGACGAGCGCAACAGCTGCGTCAGC 7228
      : : : : : ||| ||| : : : : : |||
1230 ..... 1230
7229 ACCCGCAGAGAAATTCCTGAGAGAGAACAGCTGCTCCAGAGAGGAG 7278
      : : : : : ||| ||| : : : : : |||
1241 GlnGlnAsnTyrSerProGlnGlnGlnGlnSerAlaLysGlnAsnPr 1257
      : : : : : ||| ||| : : : : : |||
7279 GAACAGCAGCTGCCGCCCAAGAGCTGACAGAAATTCGCCGAGAGAGA 7328
      : : : : : ||| ||| : : : : : |||
1257 cArgTyrSerSerTyrGlnGlnGlnSerArgAsnGlnTyrLeuGlnGln 1274
      : : : : : ||| ||| : : : : : |||
7329 ACAGCAGCTCCGCCCTGACAGAGAGAGAAATTCCTC..... 7368
      : : : : : ||| ||| : : : : : |||
1274 LysPheAsnAlaArgValMetLeuGlnTyrGlnGlnLeuArgGlnGln 1290
      : : : : : ||| ||| : : : : : |||
7369 ..... 7369
1291 GlnArgArgLys.....GlnGlnGlnMetLysLysGlnProProse 1304
      : : : : : ||| ||| : : : : : |||
7399 CTGGAGAGAGAAATTCCTGAGAGAGAACAGCTGCCCAACAAACGGAGC 7448
      : : : : : ||| ||| : : : : : |||
1304 tGlnLysProSerAsnTyrAspSerTyrLysValGlnAspProSerT 1321
      : : : : : ||| ||| : : : : : |||
7449 AGAGCAGCTGCCGCCCAAGAGACCTACAGAAATTCCTAGAGAGAGAGC 7498
      : : : : : ||| ||| : : : : : |||
1321 yTAlaProProLysGlnProPheArgGlnAspValProProSerProSer 1337
      : : : : : ||| ||| : : : : : |||
7499 AGCTCCCTCCGGAA.....AGGAGAGAA..... 7521
      : : : : : ||| ||| : : : : : |||
1338 GlnValAlaArgLeuAsnArgLeuGlnTyrProGlnLysGlnArgProPh 1354
      : : : : : ||| ||| : : : : : |||
7522 .....CAGCAGCTGCCGCCCGCCAGAGAGCGGAGAGAAATTT 7556
      : : : : : ||| ||| : : : : : |||
1354 e 1354
7557 C 7557
seq_name: /cgm2_6/ptodata/2/1ns/5B_COMB.seq:US-08-800-644-93
seq_documentation_block:
: Sequence 93, Application US/08800644
: Patent No. 5958752
: GENERAL INFORMATION:
: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: APPLICANT: Park, Sang-Chul
: TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
: NMBR OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Knodde, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.

```

```

      ZIP: 92660
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patent Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/800,644
      FILING DATE: 14-FEB-1997
      CLASSIFICATION: 424
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/056,200
      FILING DATE: 30-APR-1993
      ATTORNEY/AGENT INFORMATION:
      NAME: Fredrick, Michael F.
      REGISTRATION NUMBER: 36,799
      REFERENCE/DOCKET NUMBER: NIH054,001A
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (714) 760-0404
      TELEFAX: (714) 760-9502
      INFORMATION FOR SEQ ID NO:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 951 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      HYPOTHEetical: NO
      ANTI-SENSE: NO
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 1507..1644
      FEATURE:
      NAME/KEY: Intron
      LOCATION: 1645..2511
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 2512..8070
      US-08-800-644-93

alignment_scores:
      quality: 216.00      length: 1219
      Ratio: 0.391      Gaps: 50
      Percent Similarity: 45.283      Percent Identity: 19.688

alignment_block:
US-09-757-781-2 x US-08-800-644-93 ..
Align seg 1/1 to: US-08-800-644-93 from: 1 to: 951

304 LysArgLeuGlnLysGlnGlnLysAlaGlnGlnLeuPheArgGln 320
      : : : : : ||| ||| : : : : : |||
4426 GAGAGCTCGAGCGCGCTGAGAGCGCCGACATGAGAGAGAGCGCGCA 4475
      : : : : : ||| ||| : : : : : |||
320 uAsnAspCysIleValArgIleAsnAspLysPheArgAsnArgArgP 337
      : : : : : ||| ||| : : : : : |||
4476 GCAG.....GAGCTAGCTGAGGAGAGC 4498
      : : : : : ||| ||| : : : : : |||
337 heGlnGlnAlaGlnIleMet.PheArgGlnAlaMetArgTyrProIleT 353
      : : : : : ||| ||| : : : : : |||
4499 AGGAGAGCGCCCGAGCGGATTAGAGCGCGCATCCGAAAGTGCAGTGC 4548
      : : : : : ||| ||| : : : : : |||
353 eTrrPheHisValProAlaAlaAsnLys...GlnGlnTyrGlnGln 369
      : : : : : ||| ||| : : : : : |||
4549 CAGCTGAAAGCGAGCGCGAGCAGCAGCGCAAGCAAGCTTACTCGAGGC 4598
      : : : : : ||| ||| : : : : : |||
369 euSerGlnSerGlnLysAsnAsnTyrTyrSerArgPheSerProAsp 385
      : : : : : ||| ||| : : : : : |||
4599 CCGGCAAGCAGAGAGCGAGAGCGCGCGCAAGCAGAGAGAAAGAGGC 4648
      : : : : : ||| ||| : : : : : |||
386 SerGlnTyrIle.....As 390
      : : : : : ||| ||| : : : : : |||

```



```

927 .....AlaAlaIleAspLysSerTyrAspLysProAlaValas 939
6124 TTCCAGAGAGAAACAGCTGAGCCAGAGAAAGAAAGAAAGCCCAAG 6173
939 pAspAsp.....AspGluGluMetGluThrLeuGluGluAspThr 953
6174 GCGGAGACAGAAATCCCAAGAGAAAGAAAGCTGAGAGAGAAAG 6223
953 Lu.....GluSerSerArgSerGlyArgGluSerVal 963
6224 AAGAGAGAGACCCCTCAAGAGACAGAAATTCGCGAGAGAGAA 6273
964 SerThrAlaSerAspGluProSerHisSerLeuGluArgGluMetAsnGlu 980
6274 CACCTGCTCCAGAGAAAGAGAGAAAGAGAGAGAGAGAGAGAGAG 6323
980 Y.....AsnGluGluGlyAspLysThr 988
6324 CAGAAATTCGCGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAG 6373
988 hrAspArgLysAspLysThrGlyGluLysLysLysAspArgAsp 1004
6374 TCCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6423
1005 LysGluLysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeuGlu 1021
6424 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6455
1021 YAspMetPheArgPheGlyLysHisArgLys...AspAspLysIleGlu 1037
6456 CGAGCAACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6505
1037 YThrGlyLysIleLysIleGluGluSerPheThrSerGluGluGluArg 1053
6506 AGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6555
1054 IleArg.....MetLysGluGluGluGluGlu 1062
6556 GTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1062
1062 gTleGlnAlaLysThrArgGluPheArgGlu..... 1072
6606 GCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1073
1073 .....ArgGlnAlaArgGluArgAspLysPheVal 1081
6656 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
1082 GluIleGlnAspPheHisArgThrPheGlyCysAspAspGluLeuMet 1098
6706 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1098
1098 rGlyGlyValSerSerTyrGluGlySerMetAla...LeuAsnAlaArg 1114
6756 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1114
1114 rGlnSerProArgGluGluGluHisMetMetAlaLeuTyrAlaGluVal 1130
6806 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1130
1131 LysLysProArgAsnSerLysPheProValAspSerAsnArgSer 1147
6856 CAGCTCCCGCCG.....CAGAGAGAGAGAGAGAGAGAGAG 1147
1147 rProSerAsnHisAspArgIleGlnArgLeuArgGluGluPheGln... 1162
6882 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1162
1163 ..GlnAlaLysGlnAspGluAspValGluAspArgArgThrThrThr 1178
6932 GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1178
1179 PheGluGlnProThrProAsnAlaArgProAlaThrGlnSerGlyArg 1195

```

```

seq_name: /cgn2_6/Plodata/2/lna/5A_COMB.seq:US-08-769-309A-4
seq.document_block:
; Sequence 4, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.,
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/769,309A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5741890and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 6605 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 192..5531
US-08-769-309A-4

alignment_scores:
Quality: 207.00 Length: 1183
Ratio: 0.422 Gaps: 50
Percent Similarity: 41.420 Percent Identity: 19.104

alignment_block:
US-09-757-781-2 x US-08-769-309A-4 ..

Align seg 1/1 to: US-08-769-309A-4 from: 1 to: 6605

```
397 AAGAGTGGAGACAGAGACTGTGAAGATGTGAGCCGAAGAGACTCCGATATAA 557
508 AGGTGGAGACAGAGACTGTGAAGATGTGAGCCGAAGAGACTCCGATATAA 557
536 MetGluGlyThrValSerLeuValPheArgGlnGluSerPheIle 552
558 GAGATGGCTACTAAGTACGCGGTGTTCACGACATCCAGATGTGGCCA 607
552 sProArgGluLeuAsnAlaGlu.....ProSer..... 561
608 GGAGGAGAACCGAATATCGAACAGATTCCTTTCAGAAACAAATTTAG 657
562 ..GlnMetGlnIleProGlySerGluThrLysAlaGluAsp..... 573
658 AAGAGCTAACCAACCCACGTAGTCCAGCGCTAATGATTTGGATTAG 707
574 .....GluAspIleVal 577
708 AAGGTGTTTAAGTTTGTGGCTTTAAATTACACTGTGAAAGGATTAAGAC 757
577 IleuThrProAspGlyThrArgGluPheLeuThrPheGluValProLeuA 594
758 AGAAGAACCTGAC...ACTGTCCAGCTACTACTGTGTAAGAAGAT...G 801
594 snAspSerGlySerAlaGlyLeuGlyValSerValLysGlyAsnArgSer 610
802 AAGGGAGGAGACGACGAGGGGCTGGC..... 827
611 LysGluAsnHisAlaAspLeuGlyIlePheValLysSerIleIleAsnG 627
828 .....GACACCCAGGAC.....CCACGCTTGGGGCTGG 856
627 yGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAspIleuIleA 644
857 AGAAGACAGATCCAAA..... 872
644 IValAlaAsnGlyGluSerLeuLeuGlyLysThrAsnGlnAspAlaMetG 660
873 .....GAAGCGAACCCCAACATCTACAGAGAAACCCGAAGAG 911
661 ThrLeuArgArgSerMetSer.....ThrGluG 670
912 ACCCTGAAGCGTGAAGCAAGACGACGCAAAATTTCTCCCGACCGGATC 961
670 yAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgIleSerLys 687
962 T.....GCCAAGCAGTGAAGGAT 981
687 yAsnGlu.....Leu 690
982 GCAAGAGAGAAAGAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1031
691 LysSerProGlySerProProGlyLProGluLeu..... 701
1032 GAATCTCGACTAGTGTCCGTGACCAGTGAACAGATCAACCTTCAAAAA 1081
701 ..... 701
1082 ATTCTTCACTCAAGGTGGGCGGCGCTGGCGCAAAAGACCAAGTTTCAGA 1131
701 ..... 701
1132 AGCCGAAGAGATGAAGTGAAGCTTCAGAGAAAGAAAGAAAGAAAGAAAG 1181
702 .....ProIleGluThrAlaLeuAspAspArgGluArgGlyIleSerH 716
1182 CCAGAAAAAAGTAGACACAGAAAGAAAGCAGAAAGCAGAGGTGCTCGA 1231
716 sSerLeuTyrSerGlyIleGlnGlyLeuAspGluSerProSerArgAsnA 733
1232 GAAACTGACGCGCTCCGAGCAAGCCACCCACAGAGCGCGGCAAGAAAGT 1281
733 IAlaIleuSerArgIleMetGlyGluSerGlyLysTyrGlnLeu..... 747
```

```

1282 CCCACGACGCCCGTTATCAGCTGAATATAGAGAAAGTTGAGCTGCCCTCA 1331
748 ..... SerProthThrValAsnMetProGlnAs 756
1332 GAGGACAGTCACTGCGTCCGAGGACCTTCTGAAAGAAAGAACTGCTCC 1381
756 pAspThrValIleIleGluAspAspArgLeuPro..... 767
1382 GTTGGGACAGAAAGTTGTGATGAGAAATATGAAGTCCACCAAGAAAGG 1431
768 ..ValLeuProProHisLeuSer.....AspGlnSer 777
1432 TTGTGGCCGAGTCACTGAGCAACCTGCGAGAGAGAACCCGAGAGAG 1481
778 SerSerSerSerHisAspAspValGlyPheValThrAlaAsp..... 791
1482 AAAACGGAGGTGGAAAGAACAGCGGCTGTGCTCAGAGTGAAGAAATGGT 1531
792 .....AlaGlyThrThrPalaIleAlaIleS 801
1532 TGGAAATGATGCAGAACCTCAGAACCCGAACTGCCAAGGAGCTGTGA 1581
801 eAspSerAlaAspCys.....Ser 807
1582 AGCTCAAGAAACGTGTGTTCCGAGAGACCTTACACAGGAGACTGAC 1631
808 LeuSerProAspValAspProValLeuAlaPheGlnArgGlyPheG1 824
1632 CTCAGTCTGAT...GAGAAAGTGTCTGTCAAAACCCCGAAAGCGGTT.. 1676
824 YArgGlnSerMetSerCyluYArgThrLysGlnPheSerAspAlaSerG 841
1677 .....GTGAGTG 1683
841 IuLeuAspPheValLysThrArgLysSerLysSerMetAspLeuGlyLe 857
1684 AGGTGAATGCTGTTCATCACAGAGAGAAATGAGGTGCAG..... 1724
858 AlaAspGluThrLysLeuAsnThrValAspAspCyluLysAlaLysSerP 874
1725 .....GGAAAGTCC 1732
874 oSerArgAspValGlyProSerLeuGlyLeuLysLysSerSerSerLeuc 891
1733 ACTAAAGACGCTTTTACACAGACTGCTTAAAAAGACTTTTTCGA... 1778
891 IuSerLeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIlePro 907
1778 ..... 1778
908 PheHisArgProArgProArgIleIleArgGlyArgGlyCysAsnGluSe 924
1779 .....AMGAACAGAAAGGAAAGGAGAGAGAGAGAGAGAGATC 1819
924 rPheArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspAspA 941
1820 A.....GGGAGCACACCTCAGGTTCTCAGCGATCTCCGG 1854
941 sPAspCyluGlyMetGluThrLeuGlnGluAspThrGlnGluSerSerArg 957
1855 ACAGCCAG.....GAGAGCAAAAGGCGGAGAGCTCTGCC 1889
958 SerGlyArgGluSerValSerThrAlaSerAspClnProSerHisSerLe 974
1890 TCATCCCTCGAGAGACCCGAGAGAGATCAGTGTCTGAAAAAGGCTTAGC 1939
974 uGluArgIleMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArgu 991
1940 CGAGGTGACAGAGATGGGAAAGCTGAAGAAGAGACTTCCGATGAGAG 1989
991 ySLysAspLysThrGly..... 996
1990 AGAAAAAAGAGAAAGTGTCTCCTCGGACATCATTAAGAAAGATGTG 2039
997 ..LysCyluLysLysAspAspArgAspLysGluLysAspLysMetLysAl 1012
2040 ACGCCCAAGAAAGCGCTTTAGACGGCTTTCGAAAGATGAT..... 2078
1012 aLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArgPheGlyLysH 1029
2078 ..... 2078
1029 IAspArgLysAspAspLysIleGlu.....LysThrGlyLysLysIle 1043
2079 .....AAAGAAATGAGCTGAGACAAAGTCAAGAGCGCTACCTGTCTTCC 2123
1044 GlnGluSerPheThrSerCylu...GluGluArgIleArgMetLysGlnG1 1059
2124 ACCGAGAGCACAGCTCTGAAATGCAAGAAATGAAAGGAGAGCGTGA 2173
1059 uGlnGluArgIleGlnAlaLysThrArgLysPhe..... 1070
2174 AGAGCCAAAGCCGGAAGAAACCAAGCGAGTGAATACCTAGTATCTT 2223
1071 .....ArgGluArgGlnAlaArgGluArg 1078
2224 GCGAAGCTTTAATTGTTGTGGATCATCAAGAAAGACAGAGAGAG 2273
1079 AspTyrAlaGluIleGlnAspPheHisArgThrPheGlyCysAsp..... 1093
2274 TCCCTCTGTGATGAGAAAGGAGCCAAAGACAAATGAGAGAGACACCA 2323
1094 .....AspGluLeuMetLysGlyLysValSerSerTyrGluGlySerM 1108
2324 GAAAGCTGATGAGCCGGAAGAAACAAAGACAGCGGAGACAGCGGATCC 2373
1108 eAlaLeuAsnAlaArgProGlnSerProArgGlnLysIleMetLasp 1124
2374 TTGCT...GGTCCCAAGAACATGATCCAGGAGGGA..... 2408
1125 AlaLeuTyrAlaGlnValLysLysProArgAsnSerLysPro..... 1138
2409 .....AGTTCCTCCCGGAGCAAGC 2428
1139 .....SerProValAspSerAsnArg...SerThrProSerAsnHisAspA 1153
2429 TGGAAAGCTTACCGAAGGAGGAGCGCTTCCACCTGGAGATCAATTAAA 2478
1153 rglIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAspGluAsp 1169
2479 GGTACTCAGCCCAAGAAAAA...TCAAGTCCAAAGCTGGAAAGAGAA 2525
1170 ValGluAspArgArgArgThrTyrSerPheGlnGlnProThrProAsnAl 1186
2526 AGCGAAAGACTCCATAGCTGGGTCTGGTGTAGACATTCATCCATCAGAC 2575
1186 aArgProAlaThrGlnSerGlyArgHisSerValSerValGluValGlnM 1203
2576 TGAACCCGCTAAAGAGAAATCCGGGTCTCAATCAACAAGTTATTCCCG 2625
1203 etGlnArgGlnArgGlnGlnGluGluArgGluSerGln..... 1215
2626 GACGAGAGAGAAAGGAGCCAGATGGGAACAGAACCAAGCCCTGTGAA 2675
1215 ..... 1215
2676 GAGCCAGGCGCAACAGGCGCAACGAAGATGACTGTGATGCCGGCGGT 2725
1216 .....GlnAlaG 1218
2726 GGTCCCTGTGTGATATGATGTGTGAAAGGAGAGAAATGAGAGGAC 2775
1218 IuArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSerSer 1234
2776 AGCAAGCCCAAAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2825

```




54 CGGCCACATGAAAGTTTTCAGCCTCATTCACGACGCGGTGACCCCGCTACC 103
35 rglYsAlaIleAlaIyAspProAsnTyrTrpIleGlnValHisArgIeu 51
|||||
104 GGAGGCCATCGCCAAAGATCCAAACTACTGATACAGGTGCATCGCTTG 153
52 GlnHisGlyAspGlyIylIeLeuAspLeuAspAspIleLeuGlyAspYa 68
|||||
154 GAACATGAGATGAGGAATACTAGACCTTGATGACATTCCTTGTGATGT 203
68 lAlaAspAspLyAspArgLeuValAlaValPheAspGluGlnAspProH 85
|||||
204 AGCAGCGATTAAGACAGACTGGTACAGTGGTTGATGAGCAGAGATCCAC 253
85 lAsHisGlyIyAspGlyThrSerAlaSerSerThrGlyThrGlnSerPro 101
|||||
254 ATCAGCGAGGTGATGGCACCAAGTCCAGGTACCGGATACCCGAGACCCA 303
102 GlnIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheGlnPr 118
|||||
304 GAGATATTGGTAGTGAGCTGGCACCAACAATGCTCAGCCTTTCAGCC 353
118 oTyrGlnAlaThrSerGluIleGlnValThrProSerValIleuArgAla 135
|||||
354 TTACCAAGCAACAAGTGAATGAGGTCAACCTTCAGTCCCTTCGAGCAA 403
135 snMetProLeuHisValArgArgSerSerAspProAlaLeuIleGlyLeu 151
|||||
404 AATGGCTTTCATGTTGCAGCGAGTAGTACCCACCTCTAATTTGGCTGC 453
152 SerThrSerValSerAspSerAsnPheSerSerGluGlnProSerArgly 168
|||||
454 TCCACTTCTGTCAGTGATGATATTTTCTCTGGAAGACCTTCAGAGAA 503
168 sasnProThrArgTrpSerThrThralaGlyPheLeuLysGlnAsnThra 185
|||||
504 AATCCCAACACCTGGTCAACAACACCTGGCTTCCTCAAGCAACAACAG 553
185 lAgLySerProLysThrCysAspArgLyAspGluAsnTyrArgSer 201
|||||
554 CTGGGAGTCTTAAACCTGCGACAGAGAG..... 582
202 LeuProAlaAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAsnAl 218
582 582
218 aArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuGln 235
582 582
235 ysgInGluGlnAspGluAspGlyThrGluGlnAspAsnSerArgValGln 251
|||||
583 GATGAGATGGGAGCAGAAAGAGATTAACAGTCGTGTGA 621
252 ProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSerIe 268
|||||
622 CCGTGGACATGCTGACACGGGTTTGGAGCATATACCAACTTTTCTCT 671
268 uAspAspMetValLysLeuValGluValProAsnAspGlyLysIleProLeuG 285
|||||
672 GATGATATAGGTAAAGCTGTAGAAAGTCCCAACATGAGGAGGCGCTCTGG 721
285 lYlIleHisValValProPheSerAlaArgGlyArgThrLeuGlyLeu 301
|||||
722 GAATCCATGTAGTGCCTTTCAGTGTCCGAGGCGGCAAGAACCTGGGGCTTA 771
302 LeuValLysArgLeuGlnLysGlyGlyLysAlaGlnHisGlnAsnLeuPh 318
|||||
772 TTAGTAAACGATTTGGAGAAAGTGTAAAGCTGAACATGAAGAAATCTTTT 821
318 eaArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnAl 335
|||||
822 TCGTGAGAAATGATTCATTTGTCAAGATTATATGATGCGCAGCTTCGAATA 871

335 rgarPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThrPro 351
|||||
872 GAAAGATTTGAAACAAGCACAACATATTTTCGCCAAGCCATGCTACACCC 921
352 IleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrGluG 368
|||||
922 ATCATTTGGTCCATGTGGTTCCTGCAGCAAAATTAAGACAGATATGACA 971
368 nLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerPro 385
|||||
972 ACTATCCAAAGTGAGAAACAATTAATTCATTCACCCCTTTTAAAGCCCTG 1021
385 spSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHisThr 401
|||||
1022 ACAGCCAGATATATTGCAACAGAGGTGTACAGTGCAGGCTTCACAG 1071
402 ValGlnArgAlaProArgLeuAsnHisProGluGlnIleAspSerH 418
|||||
1072 GTGCAGAGAGACACCCGACTGAACCCCGCTGAGCAGATAGACTTCA 1121
418 sSerArgLeuProHisSerAlaHisProSerGlyLysProProSerAlaP 435
|||||
1122 CTCGAAGACTACCTCATGAGCGCACACCCCTCGGGAAACACCATCCGCTC 1171
435 roAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGlyTyr 451
|||||
1172 CAGCGTCGGCACCTCGAATATGTTTATAGTACAGCTTAAGCAGTGGTTAT 1221
452 AsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysGlyTh 468
|||||
1222 AACACCAAAAAAATAGGCAAGAGGCTTAATATCCACCTTAAGAAAGGTAC 1271
468 rGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyLys 485
|||||
1272 AGAAGTTTGGATTTCAGATCATCTTCAGAGATGAACAATAGTGCTCT 1321
485 eAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIleGln 501
|||||
1322 CAGCTCCATTCATATGAAAAACATTTCTCCCGGGGGCGGCCCATTCAG 1371
502 AspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyValAs 518
|||||
1372 GATGCCGCACTTAAGCAGAGACAGACACTTAATAGAGTAAATGAGTAGA 1421
518 pLeuValGlyLysSerGlnGluGlnValValSerLeuLeuArgSerThrL 535
|||||
1422 TTTAGTGGGCAAAATCCCAAGAGAAAGTTGTTCGCTGTGAGAGACACCA 1471
535 yMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAlaPhe 551
|||||
1472 AGATGGAAAGAACTGTGAGCCTTGTGCTTTCGCCAGAAACAGCGCTTC 1521
552 HisProAlaGluLeuAsnAlaGluProSerGlnMetGlnIleProLysG 568
|||||
1522 CACCCAAAGGGAAGTCTT..... 1536
568 uThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrArg 585
|||||
1537 ... AAAGCAGAAAGATGAGGATATTGTTCTTAACCTGATGGACACCGAG 1582
585 lUpPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyLeu 601
|||||
1583 AATTTCGTGATTTGAAAGTCCACATTAAATGATTCAGGATCTCAGAGCTT 1632
602 GlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeuG 618
|||||
1633 GGTGTCAAGTGTCAAAAGGTAAACCGGTCAAAAGAACACACAGATTTGG 1682
618 yLlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspGly 635
|||||
1683 AATCTTTGTCAAGTCCATTTAATGAGAGAGACACATCTTAAGATGCA 1732

```

635 rgluuarlyValAsnAspGlnLeuIleAlaValAlaSnGlyGlnSerLeuLeu 651
|||||
1733 GGGTGGGGGATGATGATCACTGATGACGATTAATGAGATCCCTGTTG 1782
652 GlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSerThr 668
|||||
1783 GCGAACAACAACCAAGATGCCATGGAACCCCTAAGAAGGCTATGCTAC 1832
668 rgluGluAsnLysArgGlyMetIleGlnLeuIleAlaAlaArgArgIles 685
|||||
1833 TGAAGCAATTAACGAGGAATGATCCAGCTTATTGTTGCAAGAGATAA 1882
685 erLysCysAsnGlnLeuLysSerProGlySerProGlyProGluLeu 701
|||||
1883 GCAAGTCGCAATGAGCTGAAGTCACTGGGAGCCCCCTGGACCTGAGCTG 1932
702 ProIleGluThrAlaLeuAspAspArgGluArgIleSerHisSerIle 718
|||||
1933 CCCATTGAAACACGCTTGATGATGAGAGACGAAAGATTCCCATTCCT 1982
718 uTySerGlyIleGluGlyLeuAspGlnSerProSerArgAsnAlaAla 735
|||||
1983 CTACAGTGGGATTGAGGGCTGATGATCCCAAGCAGAAATGCTGCC 2032
735 euSerArgIleMetGlyGlnSerGlyLysTyrlinLeuSerProThrVal 751
|||||
2033 TCAGTGGGATGAATG.....GGTAAATACAGCTGTCCCTACAGTG 2073
752 AsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuProVa 768
|||||
2074 AATATGCCCCAAGATGACACTGTCATTTATGAAAGATGACAGGTGGCCAGT 2123
768 IleuProThrHisLeuSerAspGlnSerSerSerSerHisAspAspV 785
|||||
2124 GCTTCCCTCCACATCTCTGACACAGTCTCTCCAGCTCCCATGATGATG 2173
785 alGlyIleValIleThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIleSer 801
|||||
2174 TGGGGTTTGGAGCGGAGATGTGGTACTTGGGCCAAGGCTGCATCTAGT 2223
802 AspSerAlaAspCysSerLeuSerProAspValAspProValLeuAlaPh 818
|||||
2224 GATTACGCCACATGCTCTTTGAGTCCAGATGTTGATCACTTCTTGCTTT 2273
818 egluIatrgLuglyPheGlyValArgGlnSerMetSerGluLysArgThrLysG 835
|||||
2274 TCAACGAGGAAGATTGGACGCTAG..... 2298
835 InPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSerLys 851
2298 ..... 2298
852 SerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAlaSp 868
|||||
2299 .....ATAGCTGACGAGACTAAACTCAATACAGTGGATGA 2333
868 pgluLysAlaGlySerProSerArgAspValGlyProSerLeuGlyLeu 885
|||||
2334 CCAAGAAAGCGAGTTCTCCAGCAGAGATGTGGGCTTCCCTGGGCTCTGA 2383
885 yLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluValThr 901
|||||
2384 AGAAGTCACACTGCTGTGGAGATCTGCAGACCGGAGTTGCCGAGGTACT 2433
902 LeuAsnGlyAspIleProPheHisArgProArgProArgIleIleAlaG 918
|||||
2434 TTGAATGGGATATTCCTTCCATGCTCCACGCGCGCGGATTAACACAGG 2483
918 yArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrlaspl 935
|||||
2484 CAGGGGATGCAATGAGAGCTTCAGAGCTGCCATCGACAAATCTTATGATA 2533
935 yspProAlaValAspAspAspGlnLysMetGluThrLeuGluGluAsp 951
|||||
2534 AACCCGCGGTAGATGATGATGAAGCATGGAGACTTGGAGAGAC. 2583
952 ThrGluGlnSerSerArgSerGlyArgGlnSerValSerThrAlaSerAs 968
|||||
2584 ACAAGAAAGTTCAAGATCATGAGGAGAGAGCTCTATCCACAGCAGCTGA 2633
968 pgluProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGluLysG 985
|||||
2634 TCAGCTTCCCACTCTCTGGAGAGACAAATGAATGAAACCAAGAGAAAG 2683
985 LysPlyThrAspArgLysLysAspLysThrGlyLysGluLysLysLys 1001
|||||
2684 GTGATAAGCTGATGAGAAAAAGATAAACTGGAAAAAGAAAGAAAAA 2733
1002 AspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLeuL 1018
|||||
2734 GATAGAGATTAAGCAGAAAGCATTAATTAAGCCCAAGAAAGGATGCTGA 2783
1018 sgluLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspLysI 1035
|||||
2784 GGGCTTGAGACATGTTCAAGGTTGGCAAAACATCGAAAGATGACAGA 2833
1035 leGluLysThrGlyLysIleLysIleGlnGlnSerPheThrSerGluGlu 1051
|||||
2834 TTGGAAGAACGGGTAAATAAATACAGGATCCTTACATCAGAAAGAG 2883
1052 GluArgIleArgMetLysGlnGlnGluArgIleGlnAlaLysThrAr 1068
|||||
2884 GACAGCATACGATTAAGCAGACGACGACGAGAGATTCAGCCCAAACTCG 2933
1068 ggluPheArgGluArgGlnAlaArgGluArgAspTyrlaGluIleGlnA 1085
|||||
2934 AGAATTTAGGAAGCAGACACTCGAGAGCGGACTATGCTGAATTCAG 2983
1085 sprPheHisArgThrPheGlyCysAspAspGlnLeuMetTyrlGlyVal 1101
|||||
2984 ATTTTCATCGGACATTTGGCTGTGATGATGATGAATGATGAGGGGAGTT 3033
1102 SerSerTyrlGlnGlySerMetAlaLeuAsnAlaArgProGlnSerProAr 1118
|||||
3034 TCTTCTTATGAGGTTCCATGGCTCTCAACGCTGAGCTCAGAGCCACG 3083
1118 ggluGlyHisMetAspAlaLeuTyrlaGlnValLysLysProArg 1135
|||||
3084 AGAAGGCATATGATGATGATGCTTGTATGCCCAAGTCAAGAACCGCGGA 3133
1135 snSerLysProSerProValAspSerAsnArgSerThrProSerAsnHis 1151
|||||
3134 ATTCCAAACCTCAGCTGTAGACAGTAAACAGATCACTCTACGCAATCAT 3183
1152 AspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAspG 1168
|||||
3184 GATGCGATACAGCCTCTGAGCAGAAATTTACAGCAGCAAAAGCAAGATGA 3233
1168 uAspValGluAspArgArgArgThrTyrlSerPheGluGlnProThrProA 1185
|||||
3234 AGATGTAGAAGATGCTGCGCGGACCTATAGTTTGAGCAAAACCTGCGCGA 3283
1185 snAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGluVal 1201
|||||
3284 ACGCAGCGCCGCGCAGCAGAGCGGCGACACTCGTGTCCGTGGAGGTG 3333
1202 GlnMetGlnArgGlnArgGlnGlnGluArgLysSerSerGlnGlnAlaG 1218
|||||
3334 CAGATGCAAGCGGACGCGCAGAGAGAGCGCAGAGCTCCACAGCAGCCCA 3383
1218 nArgGlnTyrlSerSerLeuProArgGlnSerArgLysAsnAlaSerSerV 1235
|||||
3384 GCGCAGTACAGCTCTGCTCGCGCAAAAGCAGAAAAAATCCAGCTGG 3433
1235 alSerGlnAspSerTyrlGlnGlnAsnTyrlSerProGlyGluGlyPheGln 1251
|||||

```

```

3434 TCCTCCAGACTCTTGGAGGACGAAGTACTCCCTGGGGAAGCCTTCAG 3483
1252 SerAlaTylGluAsnProArgTyrSerSerTyrGlnGlySerArgAsnG1 1268
3484 AGTGGCAAGAGAACCCAGGACTCCAGCTACCAAGGCTCCAGAAACGG 3533
1268 yTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGlnThrGln 1285
3534 CTACTCGGAGGACATGGCTTCAACGCGAGGTATGCTGGAAACTCAGG 3583
1285 IuLeuLeuArgGlnGlnGlnArgArgTyrGlnGlnGlnMetLysLysGln 1301
3584 AGCTCTTGGCCCGAGAACAGAGCGGAGAGACACACATATGAAAGCAG 3633
1302 ProProSerGlnGlyProSerAsnTyrAspSerTyrLysLysValGlnAs 1318
3634 CCTGCTTCCGAGGGGCCAGCACTATGACTGCTATACAAAGTCCAGGA 3683
1318 ProSerTyrAlaProProLysGlyProPheArgGlnAspValProProS 1335
3684 CCCAGTACGAGCCCTCCCAAGGCGCCTCCGCAAGATGCGCCCTT 3733
1335 eProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGlnLysGly 1351
3734 CCCCCTTCTCAGGTTCGAGAGGTGAACAGACTCAGACTCCTGAGAAAGG 3783
1352 ArgProPheTyrSer 1356
3784 AGGCCCTTCTATTTCC 3798

```

seq_name: /SIDSL/gcgdata/hold-1/geneseq/geneseq-emb1/AA2001A.DAT:AAH18587

seq_documentation_block:

ID AAH18587 standard; cDNA; 2718 BP.

AAH18587:

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:18774.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000BP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu N, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 8; SEQ ID 18774; 2537PP + CD ROM; English.

The present invention describes primer sets for synthesizing 5602

full-length cDNAs defined in the specification. Where a primer set

comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2718 BP; 840 A; 586 C; 688 G; 604 T; 0 other;

alignment_scores: Length: 909
Quality: 4409.00 Gaps: 2
Ratio: 5.068
Percent Similarity: 95.710 Percent Identity: 94.829

alignment_block:

US-09-757-781-2 x AAH18587 ..

Align seg 1/1 to: AAH18587 from: 1 to: 2718

```

126 GluValThrProSerValLeuArgAlaAsnMetProLeuHisValArg 142
|||||
3 GAGTCAACACCTTCAAGTCTTCGAGCAAAATATGCTTATATTTGAGG 52
142 gSerSerAspProAlaLeuIleGlyLeuSerThrSerValSerAspSer 159
|||||
53 CAGTGTGACCCAGCTTAATTGGCCTTCACATTGCTGAGTATGTA 102
159 snPheSerSerGlnGluProSerArgLysAsnProThrArgTyrSerThr 175
|||||
103 ATTTTCTCTGAGAGAGCCTTCAGGAAATATCCACACGCTGATCAACA 152
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAs 192
|||||
153 ACAGTGTGCTTCTCAAGCAGAACACTGCTGGAGTCTTAAGCCTGCGA 202
192 PARGLysLysAspGluAsnTyrArgSerLeuProArgAspThrSerAsn 209
|||||
203 CAGGAAGAAAGATGAAACTACAGAGCCTCCGCGGATACTAGTAACT 252
209 rPheSerGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225
|||||
253 GGTCTTAACCAATTTCAAGAGCAATGCTGCGTCTGTGAGTCCAGT 302
226 HisProMetValGlyLysTyrLeuGlnGlnGlnGlnGlnGlnGlnGln 242
|||||
303 CACCAATGTGTGGGCAAGTGGCAGAGAAACAGAAAGACAGATGAGATGG 352
242 yThrGluGlnAspAsnSerArgValGluProValGlyHisAlaAspThrG 259
|||||
353 GACAGAAAGAGATACAGTGTGTGAACTGTGGCATGCTGACACGG 402
259 yLeuGlnHisIleProAsnPheSerLeuAspAspMetValLysLeuVal 275
|||||
403 GTTTGGACATATATCCCAACTTTTCTGTGATGATTAAGTAAAGCTGTA 452
276 GluValProAsnAspGlyGlyProLeuGlyTyrLeuHisValValProPhe 292
|||||
453 GAATGCCCAAGATGAGGAGCTCTGGAATTCATATGAGCTTTGAG 502

```

292 rAlaAaGgLyGlyArgThrLeuGlyLeuValLysArgLeuGluysG 309
 |||||
 503 TGCTCAGCGCGCAGAACCTGGGGTTATTAGTAAAAACGATGGAGAAAG 552
 309 LyGlyLysAlaGluHisGluAsnLeuPheArgGluAsnAspCysIleVal 325
 |||||
 553 GTGGTAAAGCTGAACATGAAAATCTTTTGTGTGAGAAATGATTCATGTGTC 602
 326 ArgILAsnAspGlyAspLeuArgAsnArgPheGluGlnIleVal 342
 |||||
 603 AGGATTATATATGCGACCTTCGAAATAGAAATTTGAACAGACACACA 652
 342 sMetPheArgGlnAlaMetArgThrProIleIleTrpPheHisValValP 359
 |||||
 653 TATGTTTCGCCAAGCCATGCGTACACCATCATTTGGTTCATGTGTTTC 702
 359 rAlaAlaAsnLysGluGlnTrpGluGlnLeuSerGlnSerGluLysAsn 375
 |||||
 703 CTTCAGCAAAATTAAGACAGTATGAACTATCCCAAAAGTGAAAGAAC 752
 376 AspTyrTyrSerSerArgPheSerProAspSerGlnTyrIleAspAsnAr 392
 |||||
 753 AATTACTATTCAGCCGTTTACCCCTGACAGCAGCATATTTGACAAACAG 802
 392 gSerValAsnSerAlaGlyLeuHisThrValGlnArgAlaProArgLeuA 409
 |||||
 803 GAGTGTGAACAGTGCAGGCGCTCACACGCTGCAGAGACACCCGACGTGA 852
 409 snHisProProGluGlnIleAspSerHisSerArgLeuProHisSerAla 425
 |||||
 853 ACCACCGCTGAGCAGATAGACTCTACTCAAGACTCACTCAAGGCGCA 902
 426 HisProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsnVa 442
 |||||
 903 CACCCCTCGGGAAACACCATCGCGTCCAGCCTCGGCACCTCAGATGT 952
 442 lPheSerThrThrValSerSerGlyTyrAsnThrLysLysIleGlyLysA 459
 |||||
 953 ATTAGTACACACTGTAAGCAGTGTATTACACCAAAAAAATAGGCAAGA 1002
 459 rGluAsnAsnIleGlnLeuLysGlyThrGluGlyLeuGlyPheSerIle 475
 |||||
 1003 GGCTTATATTCAGCTTAAGAAAGTACAGAAAGTTTGGGATTCAGCATC 1052
 476 ThrSerArgAspValThrIleGlyGlySerAlaProIleTyrValLysAs 492
 |||||
 1053 ACTTCCAGAGATGTAAACAATAGTGGCTCAGCTCCATCTATGTGAAAAA 1102
 492 nIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuLysAlaGlyA 509
 |||||
 1103 CATTCCTCCCGGGGGCGGCCCATTCAGATGCCGACTTAAGGCAGAGAG 1152
 509 sPheArgLeuIleGluValAsnGlyValAspLeuValGlyLysSerGlnGlu 525
 |||||
 1153 ACAGACTTATAGAGGTAAATGAGTATTAAGGGCAAAATCCCAAGAG 1202
 526 GluValValSerLeuLeuArgSerThrLysMetGluGlyThrValSerIle 542
 |||||
 1203 GAAGTGTGTTTCGCTTGAGAGACCAACAAGATGAAGAAGTGTGAGACT 1252
 542 uLeuValPheArgGlnGluAspAlaPheHisProArgGluLeuAsnAlaG 559
 |||||
 1253 TCGGTCTTTTCGCCAGAAAGACGCTTCCACCCAAAGGAACTGAATGACAG 1302
 559 lProSerGlnMetGlnIleProLysGluThrLysAlaGluAspGluAsp 575
 |||||
 1303 AGCCAAAGCCAGATGCAGATTCCAAAAGAAAGCAAGAAAGATGAGAGAT 1352
 576 lIleValLeuThrProAspGlyThrArgGluPheLeuThrPheGluValPr 592
 |||||
 1353 ATTGTCTTACACCTGATGGCACCAAGGAATTTCTGACATTTTGAAGTCCC 1402

592 OlEuAsnAspSerGlySerAlaGlyLeuGlyValSerValLysGlyAsnA 609
 |||||
 1403 ACTTACTGATTCACAGATCTGCAGCGCTTGGTGCAGTGCATAAAGTAAAC 1452
 609 rGSerLysGluAsnHisAlaAspLeuGlyIlePheValLysSerIleIle 625
 |||||
 1453 GGTCAAAAGAGAACACGACGATTTGGAAATCTTTGTCAAGTTCATTAAT 1502
 626 AsnGlyGlyAlaAlaSerLysAspGlyArgLeuArgValAsnAspGlnLe 642
 |||||
 1503 AATGAGAGACAGCATCTTAAAGATGGAAGCCTTGGGTGAATGATCAACT 1552
 642 uIleAlaValAsnGlyLysLeuSerLeuGlyLysThrAsnGlnAspAlaM 659
 |||||
 1553 GATAGCAGTAAATGAGAAATCCCTGTTGGCAAGACAAACCAAGATGCCA 1602
 659 eGluThrLeuArgArgSerMetSerThrGluGlyAsnLysArgGlyMet 675
 |||||
 1603 TGGAAACCTTAAGAAAGTCTATGTCTACTGAAGGCAATAAACGAGAAATG 1652
 676 lIleGlnLeuIleValAlaArgArgIleSerLysCysAsnGluLeuLysSe 692
 |||||
 1653 ATCCAGCTTATTTGTTGCCAAGAGAAATAAAGCAAGTGCATAGCTGAAGTC 1702
 692 rProGlySerProProGlyProGluLeuProIleGluThrAlaLeuAspA 709
 |||||
 1703 ACCTGGAGAGCCCGCTGACCTGACCTGACCTGAAACACGCTTGATG 1752
 709 sPheArgLysArgArgIleSerHisSerLeuLysArgGlyIleGluGlyLeu 725
 |||||
 1753 ATAGAGAAACGAAGAAATTTCCATTCCCTTACAGTGGAGTGAAGGGGCTT 1802
 726 AspGluSerProSerArgAsnAlaAlaLeuSerArgIleMetGlyLysSe 742
 |||||
 1803 GATGAATTCGCCACAGCAAGAAATGCTGCCCTCAGTAGATGAATAG..... 1844
 742 rGlyLysTyrGlnLeuSerProThrValAsnMetProGlnAspAspThrV 759
 |||||
 1845 .GGTAAATACACAGCTGTCCCTACAGTGAATATGCCCAAGATGACACTG 1893
 759 aIleIleGluAspAspArgLeuProValLeuProProHisLeuSerAsp 775
 |||||
 1894 TCATTATTAAGAGATGACAGGTGCGCAGGTCTTCCACATCTCTGTAC 1943
 776 GlnSerSerSerSerSerHisAspAspValGlyPheValThrAlaAspAl 792
 |||||
 1944 CAGTCCCTCTCCACCTCCCATGATGATGTGGGTTTGTGACGCGCAGATGC 1993
 792 aGlyThrTrpAlaLysAlaAlaIleSerAspSerAlaAspCysSerLeuS 809
 |||||
 1994 TGGTACTTGGGCGCAAGGCTGCATACAGATTCAGCCGACGTGCTTTGA 2043
 809 eProAspValAspProValLeuAlaPheGlnArgGluGlyPheGlyArg 825
 |||||
 2044 GTCCAGATGTTGATCCAGTCTTGTCTTTCAACGAAAGAAATTTGGAGGT 2093
 826 GlnSerMetSerGluLysArgThrLysGlnPheSerAspAlaSerGlnLe 842
 |||||
 2094 CAGACT..... 2099
 842 uAspPheValLysThrArgLysSerLysSerMetAspLeuGlyIleAlaA 859
 |||||
 2100G 2100
 859 sPgluThrLysLeuAsnThrValAspAspGlnLysAlaGlySerProSer 875
 |||||
 2101 ACGAGACTAAACTCAATACAGTGAATGACCAAGAAAGCGGTTCTCCACGC 2150
 876 ArgAspValGlyProSerLeuGlyLysLysSerSerSerLeuGlyLys 892
 |||||
 2151 AGAGATGTGGGTCTCTCCCTGGGTCTGAAGAAGTCAACCTATTTGGAGAG 2200
 892 rLeuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIleProPhe 909

```

|||||
2201 TCTGAGACCGCAGTTGCGAGGTGACTTAAATGGGATTCCTTCC 2250
909 1AARGProARProAR1leilearg1yArgrg1yCysAsnGluSerPhe 925
|||||
2251 ATCGTCCACGGCCGGGATTAATCAGAGGAGGGGATGCAATGAGAGCTTC 2300
926 ARGAlaAla11leasplysSerTyrAspLysProAlaValaAspAspAs 942
|||||
2301 AGAGCTGCGCATCGACAAATCTTATGATTAACCCGGGTAGATGATGATGA 2350
942 pGIuGIyMetGluThrLeuGIuGIuAspThrGluGluSerArqSerg 959
|||||
2351 TGAAGCATGAGACCTTGGAAGAACACAGAGAAGAACTTAATCAATCAG 2400
959 1yArGIuSerValSerThrAlaSerAspGlnProSerHisSerLeuGlu 975
|||||
2401 GGAGAGAGTCTGTATCCACAGCCAGTGCATCAGCTCCACATCTCGGAG 2450
976 ArgGlnMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArgLysLy 992
|||||
2451 AGACAATGATGATGAAACCAAGAGAAAGGTGATTAAGACTGATAGAAAAA 2500
992 sAspLysThrGlyLysGlyLysLysLysAspArgAspLysGlyLysAsp 1009
|||||
2501 GGATTAACCTGGAAGAGAAAGAAAGATAGATTAAGAGAGAGATGA 2550
1009 ySMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArg 1025
|||||
2551 AAATGAAGCCAGAGAGGAGATGCTGAAGGCTGGGAGACATGTTCCAGC 2600
1026 PheGlyLysHisArgLysAspAspLys 1034
2601 CTTCGCAAACTGACAGCCGAGAGAGCA 2627

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH78369
seq_documentation_block:
ID AAH78369 standard; DNA: 2718 BP.
XX
XX AAH78369;
XX
XX 26-NOV-2001 (first entry)
XX
XX Nucleotide sequence of a human protein kinase/protein phosphatase.
XX
XX Human; protein kinase; protein phosphatase; signal transduction; ss.
XX
XX Homo sapiens.
XX
XX key Location/Qualifiers
XX CDS 33..2630
XX FT /tag= "a
XX FT /product= "protein kinase/protein phosphatase"
XX
XX PN WO200109316-A1.
XX
XX PD 08-FEB-2001.
XX
XX PF 28-JUL-2000; 2000MO-IP05061.
XX
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 18-OCT-1999; 99US-0159590.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 17-FEB-2000; 2000US-0183322.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Funahashi S;
XX Senoo C, Nezu J;

```

```

XX WPI: 2001-570286/64.
DR P-PSDB; AAG67610.
XX
XX New genes encoding proteins with protein kinase/protein phosphatase
XX activity, useful in the diagnosis and treatment of diseases -
XX
XX Claim 1: Page 67-76; 233pp; Japanese.
XX
XX The present sequence encodes a human protein kinase/protein phosphatase.
XX It is expected that the protein kinase/protein phosphatase gene
XX participates in signal transduction in cells. The protein kinase/protein
XX phosphatase polypeptides and polynucleotides are useful for developing
XX diagnostics and treatment agents for human and animal diseases. The
XX protein kinase/protein phosphatase polypeptides are useful as target
XX molecules in designing novel drugs. The protein kinase/protein
XX phosphatase polynucleotides are useful as a source of probes and
XX primers, which may be used to isolate homologous sequences.
XX
XX Sequence 2718 BP: 840 A; 586 C; 688 G; 604 T; 0 other;

alignment_scores:
Quality: 4409.00 Length: 909
Ratio: 5.068 Gaps: 2
Percent Similarity: 95.710 Percent Identity: 94.829

alignment_block:
US-09-757-781-2 x AAH78369 ..
Align seg 1/1 to: AAH78369 from: 1 to: 2718

126 GluValThrProSerValLeuArgAlaAsnMetProLeuHisValaArgAr 142
3 GAGGTACACACCTTCAGTCTTCGAGCAAAATGAGCTCTTATGTTGACAG 52
142 gSerSerAspProAlaLeu1leGlyLeuSerThrSerValSerAspSera 159
|||||
53 CAGTGTGACCCAGCCTCTAATTGGCCTCTCCATCTGTCATGATGATGA 102
159 snPheSerSerGluGluProSerArqLysAsnProThrArgTyrPsrThr 175
|||||
103 ATTTTCTCTGAGAGAGGCTTCAGAGAAATCCACACGCTGATCAACA 152
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAs 192
|||||
153 ACAGCTGCTTCTCTCAGAGCAACACTGCTGGAGTCTCTAAAGCCTGGA 202
192 pArgLysLysAspGluAsnTyrArgSerLeuProArgAspThrSerAsn 209
|||||
203 CAGGAGAAAGATGAAACTACAGAAAGCTCCCGCGGATACTACTACT 252
209 rPserAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225
|||||
253 GGTCTTAACAATTTCAGAGAGCAATGCTCGCTGCTGTGAGTCCAGT 302
226 HisProMetValGlyLysTrpLeuGlyLysGlnGluAspGluAspG 242
|||||
303 CACCAATGCTGGGCAAGTGGCAGAGAACAGAACAGACAGATGAGGATG 352
242 yThrGluGluAspAsnSerArqValGluProValGlyHisAlaAspThrG 259
|||||
353 GACAGAAAGAGATTAACACTGCTGTGAACCTGTTGACATGCTGACACG 402
259 1yLeuGluHis1leProAsnPheSerLeuAspAspMetValLysLeuVal 275
|||||
403 GTTGGACATATACCAACTTTCTCTGATGATGATGTAAGGCTCGTA 452
276 GluVal1ProAsnAspGlyLysProLeuGly11leHisVala1ProPhe 292
|||||
453 GAAGTCCCAACAGATGAGGCGCTCTGGAGATCCATGATGAGCTTTACG 502
292 rAlaArgLysGlyLysArgThrLeuGlyLeuValValLysArgLeuGlyLysG 309

```

```

|||||
503 TCGTCGAGCGCGCAGAACCTGGGGTTATTAGTAAACGATTGGAGAAAG 552
309 IYGLYLSAIAGLIuHISGLuASnLeuPheArgIuASnAspCysIleVal 325
553 GTGGTAAGCTGACATGATAAAATCTTTTCGTGAGAAATGATTGCATTGTC 602
326 ArgIleASnAspGlyASpLeuArgASnArgPheGluGlnAlaGlnI 342
603 AGCATTTAAATGATGGCGACCTTCGAAATAGAAATTTGAACAAGCACA 652
342 sMetHearGlnAlaMetArgThrProIleTrrPheHISValValP 359
653 TATGTTTCGCCAAGCCATGCTACACCCATCATTTGGTTCCATGTGTTCC 702
359 roAlAlaASnLysGluGlnTYrGluGlnLeuSerGlnSerGluLysAsn 375
703 CTGCAGCAAAATAAAGACAGTATGAAACAATCTCCAAAGTGAGAAAGAC 752
376 AsnTYrTYrSerSerArgPheSerProAspSerGlnTYrIleAspAsn 392
753 AATTACTATTCAAGCCGTTTACCCCTGCACAGCCAGTATATTGACAAAC 802
392 gSerValASnSerAlaGlyLeuHISThrValGlnArgAlaProArgLeu 409
803 GAGTGTGAACAGTGCAGGGCTTCACACGGTGCAGAGACACCCGCTGA 852
409 snHISProProGluGlnIleASpSerHISerArgLeuProHISerAla 425
853 ACCACCCGCTGACGACGATAGACTCTCACTCAAGACTACCTCATAGCCGA 902
426 HIsProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsn 442
903 CACCCCTCGGAAACACACCATCCGCTCCAGCCTCGGACATCAGAAATG 952
442 lPheSerThrThrValSerSerGlyTYrAsnThrLysLysIleGlyLys 459
953 AATTAGTAGACTGTAAAGCATGTGTTATTAACACAAAAAATAGGCMAA 1002
459 rGLeuASnIleGlnLeuLysLysGlyThrGluGlyLeuGlyPheSerIle 475
1003 GCGTTAAATCCAGCTTAAGAAAGTACGAAAGGTTGGGATTCCAGCATC 1052
476 ThrSerArgAspValThrIleGlySerAlaProIleTYrValLysAs 492
1053 ACTTCCAGAGATGTAAACATAGTGGCTCAGCTCCAAATCATGTGAAAA 1102
492 nIleLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuLysAlaGly 509
1103 CATTTCTCCCGCGGGGGCGCATTCAGATGCGCATTAAGCGAGAG 1152
509 sPArgLeuIleGluValASnGlyValASpLeuValGlyLysSerGlnLys 525
1153 ACAGACTTATAGAGTAAATGAGATGATTAGTGGCAAAATCCCAAGAG 1202
526 GluValValSerLeuLeuArgSerThrLysMetGluGlyThrValSerIle 542
1203 GAAGTTGTTCCCTGTGAGAAAGCACAAGATGGAAGAACTGTGAGCCT 1252
542 uLeuValPheArgGlnGluASpAlaPheHISProArgIuLeuASnAlaG 559
1253 TCTGGCTTTCCCGAAGAGACGCTTCCACCCAAAGGAATGAATGAGAG 1302
559 lUpProSerGlnMetGlnIleProLysGluThrLysAlaGluAspGluAsp 575
1303 ACCCAAGCCAGATGCAGATTCAAAAAAGAAACGAAAGCAGAAAGTGAGAT 1352
576 lIleValLeuThrProAspGlyThrArgGluPheLeuThrPheGluVal 592
1353 AATTGTTCTTACACCTGATGACACAGGAATTTCTGACATTTGAAGTCCC 1402
592 oLeuASnAspSerGlySerAlaGlyLeuGlyValSerValLysGlyAsn 609
|||||

```

```

1403 ACTTACTGATTACAGATCTGCAGGCTTGGTGTCACTGTCAAAAGTAAACC 1452
609 rGSerLysGluASnHISAlaASpLeuGlyIlePheValLysSerIleIle 625
1453 GGTCAAAAGAGAAACACGACGATTTGGGAATCTTTGTCAAGTCCATTATT 1502
626 AsnGlyGlyAlaAlaSerLysAspGlyArgLeuArgValASnAspLysIle 642
1503 AATGAGAGAGACACATCTAAAGATGGAAGGCTTCGGGTGCAATGATCAACT 1552
642 lIleAlaValASnGlyLysSerLeuLeuGlyLysThrASnGlnAspAlaM 659
1553 GATAGCATTAATGAGAAATCCCTGTGGGCAAGACAAACCAAGATGCCA 1602
659 eGcIuThrLeuArgArgSerMetSerThrGluGlyASnLysArgGlyMet 675
1603 TGGAAACCCTAAGACAGTCTATCTCTACGTAAAGCATAAACAGAGATG 1652
676 lIleGlnLeuIleValAlaArgArgIleSerLysCysAsnGluLeuLysSe 692
1653 ATCCACCTTATTGTCCAAAGAGAAATAGCAAGTCAATGACCTGCAAGTGC 1702
692 rProGlySerProProGlyProGluLeuProIleGluThrAlaLeuAsp 709
1703 ACCTGGAGCCCTCGAGACTGAGCTGCCATTTGAACAGCTTGGATG 1752
709 sPArgGluArgArgIleSerHISerLeuTYrSerGlyIleGluGlyLeu 725
1753 ATAGAGACAGCAAAATTTCCCATTCCTCTACAGTGGGATTTGAGGGCTT 1802
726 AspGluSerProSerArgASnAlaAlaLeuSerArgIleMetGlyLys 742
1803 GATGAATCGCCCGACAGAAATGCTGCCCTCAGTAGATATG..... 1844
742 rGlyLysTYrGlnLeuSerProThrValASnMetProGlnAspAspHy 759
1845 .GGTAAATACACAGCTGCTCCCTACAGTATGATGCCCCAAGATGACACTG 1893
759 aIleIleGluASpAspArgLeuProValLeuProProHISLeuSerAsp 775
1894 TCAATTATAGAAATGACAGGTTGCCAGTCTTCTCCACATCTCTGTGAC 1943
776 GlnSerSerSerSerSerHISAspAspValGlyPheValThrAlaAsp 792
1944 CAGTCCCTCTCCAGCTCCCATGATGATGTTGGTTGTGACGACAGATGC 1993
792 aGlyThrTrrPalaLysAlaAlaIleSerAspSerAlaAspCysSerLeus 809
1994 TGGTACTTGGGCCAAGGCTCAATCAGTGAATTCAGCCGACTGCTCTTTGA 2043
809 ePProAspValAspProValLeuAlaPheGlnArgGluGlyPheGlyArg 825
2044 GTCCAGATGTGATCCAGTCTTGTCTTTTCAACGAGAAGATTTGGACGT 2093
826 GlnSerMetSerGluLysArgThrLysGlnPheSerAspAlaSerGlnLe 842
2094 CAGACT..... 2099
842 uAspPheValLysThrArgLysSerLysSerMetAspLeuGlyIleAlaAla 859
2100 .....G 2100
859 sPGLuThrLysLeuASnThrValASpAspGlnLysAlaGlySerProSer 875
2101 ACAGAGACTAAATCATATACGTGATGACCAAGAAAGAGTCTTCCACG 2150
876 ArgAspValGlyProSerLeuGlyLeuLysLysSerSerSerLeuGlnSe 892
2151 AGAGATGTGGGTCTTCCCTCGGCTGTGAAGAAAGTCAAGCTCATTTGAGAG 2200
892 rLeuGlnThrAlaValAlaGluValThrLeuASnGlyAspIleProPheH 909
2201 TCTGCAGACCGCAGATTGCCAGGTGACTTTGAATGGGATATTCTCTTCC 2250

```

```

909 1sargProargProargilelleargGlyargGlycysasnGlySerPhe 925
2251 ATGTCACAGCGCGGATATCATCAGAGCGAGGATGATGAGAGCTTC 2300
926 ATGAlaAlaIleasplysSerTyrAspLysProAlaValAspAspAsp 942
2301 AGAGCTGCGCATGCAAAATCTTATGATTAACCCGCGCTAGCATGATGATGA 2350
942 pgluglyMetGluThrLeuGluGluAspThrGluGluSerArgSerG 959
2351 TGAAGGATGAGACCTTGCAAAACACACAGAAAGTTCAAGATTCAG 2400
959 LyrargGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGlu 975
2401 GGAGAGAGTCTGTATCCACAGCCAGATGATCAGCTCCACCTCTGGAG 2450
976 ArgGluMetAsnGlyAsnGlnGlyAspLysThrAspArgGlySly 992
2451 AGACAAATGCAATGGAACCAAGAAAGGTGATMACCTGATAGAAAAAA 2500
992 sAspLysThrGlyLysGluLysLysAspArgAspLysGluLysAspL 1009
2501 GGATTAACCTGGAAGAAAGAAAGAAAGATAGACATTAAGGACAAAGATA 2550
1009 yMetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArg 1025
2551 AATGAAAGCCAAAGAGGAAATGCTGAGGCTTGAGGACATGTTCCAGC 2600
1026 pHeGlyLysHisArgLysAspAspLys 1034
2601 CTTCGCAAACTGAGAGCCGAGAAAGACA 2627

seq_name: /SIDSL/9cgcdata/hold-geneseg/geneseg-emb1/NM2000.DAT: AAD01182
seq_documentation_block:
ID AAD01182 standard; cDNA; 2554 BP.
AC AAD01182;
XX
DT 04-OCT-2000 (first entry)
XX
DE PDZ domain-comprising PHIP (ephrin Interacting Protein) complex cDNA.
XX
KW PDZ domain; B class ephrin; Eph receptor tyrosine kinase; RTK; modulator;
cellular process; nerve cell interaction; regeneration of nerve cell;
axonogenesis; antiapoptotic; proliferative disorder; treatment; GRIP;
differentiative disorder; PHIP; ephrin interacting protein; synenin; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 306..2540
FT /*tag- a
FT /product- "PHIP (ephrin Interacting Protein) complex"
FT /note- "Homologous to C. elegans PAR-3 protein"
XX
PN W0200031124-A2.
XX
PD 02-JUN-2000.
XX
PE 19-NOV-1999; 99WO-CA01101.
XX
PR 20-NOV-1998; 98US-0109158.
XX
PA (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI Lin D, Pawson A;
XX
DR MPI; 2000-400038/34.
XX
DR P-PSDB; AATY1410.
XX
PT Isolated complex for treating proliferative or differentiative

```

```

PT disorders comprises B class ephrin and PDZ domain containing protein -
XX
PS Disclosure; Page 57-59; 59pp; English.
XX
CC The patent discloses a complex comprising of a B class ephrin and PDZ
CC domain containing protein. B class ephrins function as ligands for Eph
CC receptor tyrosine kinases (RTK) and possess a transmembrane element and a
CC highly conserved cytoplasmic tail at the C-terminus, that contains a PDZ
CC binding site. This complex is used in methods to modulate the interaction
CC of a B class ephrin and PDZ domain containing protein and to identify
CC like, axonogenesis, nerve cell interactions and regeneration of nerve
CC cells. The complex is also useful for treating proliferative or
CC differentiative disorders associated with this protein complex.
CC The present sequence is the cDNA clone encoding PHIP (ephrin Interacting
CC Protein) complex, comprising three PDZ domains. This clone is isolated
CC by screening a lambdaex10.5 day mouse embryo expression library.
CC PHIP is closely related to Caenorhabditis elegans PAR-3 protein involved
CC in regulating polarity of the early embryo. PHIP cDNA fragment was used
CC as a probe, to isolate other transcripts like GRIP and synenin proteins
CC from 10.5 day mouse embryo library.
XX
SQ Sequence 2554 BP; 650 A; 697 C; 745 G; 462 T; 0 other;

alignment_scores:
Quality: 3527.00 Length: 743
Ratio: 4.892 Gaps: 0
Percent Similarity: 97.039 Percent Identity: 91.386

alignment_block:
US-09-757-781-2 x AAD01182 ..

Align seg 1/1 to: AAD01182 from: 1 to: 2554

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||
306 ATGAAAGTGAACCGTGTGCTCGGAGGACCCGGGTGGTGGCGCGGG 355
17 YAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
|||||
356 AGATGCGCCGATGAAGATTTTCAGCTTCATCCACAGCGGCTGACCCGCT 405
34 YrArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
406 ACCGGAAGCGCGTGGCAAGATTCCAAACTGATGATACAGTCAATCC 455
51 LeuGluHisGlyAspGlyLysIleLeuAspLeuAspIleLeuCysAs 67
|||||
456 TTGGAGCATGAGATGAGAGGATTCCTAGACCTCGATGACATCCCTGTA 505
67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
|||||
506 CGTTGCTGATGACAAAGACAGACTGTGATGATTTGATGAAGAGGATC 555
84 rHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
556 CCCACCATGAGGAGATGTACACAGCGCCAGCTCCACGGAACCCAGAT 605
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
|||||
606 CCAGAGATATTCCGACATGAGCTGGGCAACACAAATGTTCTGCTTTTCA 655
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
|||||
656 CCCTTATCAAGCCCAAGTGAATTTAGAGTACGCTTCAGTTCTTGGG 705
134 LaasnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
|||||
706 CAAATATGCTCTTCATGTCGCGGAGGAGCAGCAGCTTAATCTGGC 755
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
|||||

```

756 CTTCCACTTCTGTCAGTGATTAACAACCTTTCCTCAGAGAGAGCCCTCCAG 805
 167 glysasnProthrArgTyrSerThrAlaGlyPheLeuLysGlnasn 184
 806 GAAAAACCCACCGCGTGGTCCAGCAGACAGTGGCTTTCTCAACAGACAGA 855
 184 hAlaGlySerProLysThrCysAspArgLysLysAspGluasnTyrArg 200
 856 CCCTGGAACTCCCAAAACCTGCAGACAGAAAGATGAAAACCTACAGA 905
 201 SerLeuProArgSprHisSerAsnTyrSerAsnGlnPheGlnArgAsp 217
 906 AGCTTTCACAGGGATCCCACTACTGGTCCACACAGTTCACAGCAGACAA 955
 217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTyrLeuc 234
 956 CGCCCGCTCTCCCTGAGCGCCAGCCCAATGGTAGACCGGTGGTGG 1005
 234 LuLysGlnGluAspGluAspGlyThrGluGluAspAsnSerArgVal 250
 1006 AGAAGCAGAAACAGATGAGAGAGCCACAGAAAGACACAGCCGAGTG 1055
 251 GluProValGlyHisAlaSprThrGlyLeuGluHisTrieProAsnPhe 267
 1056 GAGCCGGTTGGACATGCTGATACCGGATGGAGAACATGCCCACTTTC 1105
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
 1106 CCTCGATGATATGTAAAGCTCGTACAAAGTCCCAACGATGGAGGGCCC 1155
 284 euGlyLLeuHisValProPheSerAlaArgGlyLysArgThrLeuGly 300
 1156 TGGGAATCCATGTAGTGCCTTTCAGTCTCGAGCGCGCAGAACATTGGG 1205
 301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLe 317
 1206 TTGTATGTAGAGCGGTGGAGAAAGCGGTAAAGCTGAGCAAGAAACCT 1255
 317 upheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
 1256 TTTTCATGAGATGATGCTGATGTGAGAGATTAAAGATGAGATCTTGAA 1305
 334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
 1306 ACAGAAGATTGTGAGCAAGCAACATATGTTCCGCCAAGCTATGCGTGC 1355
 351 ProIleLeuPheHisValProAlaAlaAsnLysGluGlnTyrGly 367
 1356 CGGTGCTTTGGTTCATGTGTGTCCTCGACGAAACAGAGCAATATGA 1405
 367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
 1406 ACAACGTGCCCAACGCGAAGAAACAATCTCCCGCCGCGCTTGAGCC 1455
 384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
 1456 CTGACAGGCCACTGTGTGGCCACAGAGAGTGTGGCAACATGGCCCTCAA 1505
 401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
 1506 GCATTGGCCAGACACCCAGACTGAGTCAGACACCCAGACAGCGGTGC 1555
 417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
 1556 TCACCCCGCACTACCTCATAGTGTCCAGCCTCAACCAACACACCCGAG 1605
 434 lAlaProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
 1606 CCCCAGGCTTGGCTCCACCCAGTGTCTAGTACCAACGTAGGACAGTGTG 1655
 451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysGly 467
 1656 TACAAACGAAAGAAAGTAGGCAAGGCTCAACATCCAGCTTAAGAAAGG 1705

467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
 1706 TACAGAAAGACTGGGATTCAGACATCACCTCCCGGAGCTCACCATAGTGTG 1755
 484 lYSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
 1756 GCTCAGCTCCATTTATGTCAAGATATCTTCTCCACAGGGGCTGCAT 1805
 501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517
 1806 CAGATGGCACTCAAGCAGAGACCGGCTAAATAGAGTCATAGAGT 1855
 517 lAspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 534
 1856 AGATTATGACAGGCAATCCAGAGAAAGTGTTCCTCGTTGGAAGACA 1905
 534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnLysAla 550
 1906 CCAAGATGAGGGGAGCTGTAGCCTTCTGTCTTCGTCAGAGAAAGGCT 1955
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProly 567
 1956 TTCCACCCCAAGGAAATGATGCTGAACCAACCAAGATGCAGACTCCAA 2005
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
 2006 AGAAACCAAGCTGAGATGAGAGCTGTTCACACCCAGTGGTACCA 2055
 584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
 2056 GGGAGTTCTCATTGTAAGTTCACCTGATGATGACTCAGGATGTGCAGGG 2105
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
 2106 CTGTGTGTCACTGTCAAGGGGACCGTTCCAAAGAAACCCAGATTT 2155
 617 uGlyLLeuPheValLysSerIleIleAsnGlyLysAlaAlaSerLysAsp 634
 2156 GCGGATCTTCGTAAATCCATTTCAATGCTGAGAGCTCATCTAAAGATG 2205
 634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
 2206 GAAAGCTGAGGCTAAATGACACAGCTGATGCTGTAAGAGAAATCTCTA 2255
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
 2256 CTGGGCAAAAGCCAAACGAAAGCCATGAGACTCTACGAGAGTCCATGTC 2305
 667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
 2306 CACGAGGCGCAACAGCGTGCATGATCCACGCTCATTTGTGCGAGGCGGA 2355
 684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyLysPro 700
 2356 TCACACAGATGTAAAGAGCTGTGGTCTCTGTGGAGCCCTGTCACCTTAG 2405
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
 2406 CTGGCCATCGAGACAACTGGATGACCGAGAACGAGAGATCTCAKATC 2455
 2456 CCTTACAGTGTGAGTTCGAGGGGTGAGTGTGCGCCACAGAAACGAG 2505
 734 lAlaLeuSerArgIleMetGlyLysSerGly 743
 2506 CACTGAGCAGATATATGGTGAATCAGGA 2534
 seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-embL/NA2001A.DAT:AAH72793
 seq_documentation_block:
 ID AAH72793 standard; cDNA; 1800 BP.

```
XX AAH72793;
AC
AD 19-SEP-2001 (first entry)
DE Human cervical cancer marker nucleic acid 4067.
KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX Homo sapiens.
XX WO200142467-A2.
PN 14-JUN-2001.
PD
PE 08-DEC-2000; 2000WO-US33312.
XX 08-DEC-1999; 99US-0169681.
PR 21-DEC-1999; 99US-0171350.
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
PA (MIL-) MILENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
PS Claim 1; Page 855-856; 1051pp; English.
XX
XX The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC to select and/or assess the efficacy of a premalignant condition in a patient;
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 1800 BP; 578 A; 359 C; 466 G; 394 T; 3 other;

alignment_scores:
Quality: 2778.50 Length: 563
Ratio: 4.997 Gaps: 1
Percent Similarity: 98.757 Percent Identity: 97.869

alignment_block:
US-09-757-781-2 x AAH72793 ..
Align seg 1/1 to: AAH72793 from: 1 to: 1800

472 GlyPheSerIleThrSerArgAspValThrIleGlyGlySerAlaProIle 488
|||||
4 GGATTCAGCATCATTCCAGAGATGTAAACATATGATGGCTCAGCTCCGAT 53
488 eTyrValIysAsnIleLeuProArgGlyAlaAlaIleGlnAspGlyArgL 505
|||||
54 CTAATGAAACATTCCTCCCGGGGGCGGCATTCAGATGGCGCGAC 103
54 eulysAlaIleAspArgLeuIleGlyAlaAsnGlyAlaAspLeuValGly 521
|||||
104 TTAAGGCGAGAGACACTTATAGAGTAATGAGTAGATTAGTGGGC 153
522 LysSerGlnGluGluValSerLeuLeuArgSerThrLysMetGluG 538
|||||
154 AATTCCTCAAGAGAGAGTTGTTCCGCTGTGAGACACCAAGATGSAAG 203
```

```
538 yThrValSerLeuLeuValPheArgGlnGluAspAlaPheHisProArg 555
|||||
204 AACTGTGAGCCCTTCTGCTCTTCCGCGAGAAAGCCCTTCCACCAAGGG 253
555 luleuAsnAlaIleuProSerGlnMetGlnIleProLysGluThrLysAla 571
|||||
254 AACTGAATGCAGAGCCAAAGCCAGATGCAGATTCCAAAGAAACGAAAGCA 303
572 GluAspGluAspIleValIleuThrProAspGlyThrArgGluPheLeu 588
|||||
304 GAAGATGAGAGATATTGTTCTTACACCTGATGSCACAGGAATTTCTGAC 353
588 rPheGluValProLeuAsnAspSerGlySerAlaGlyLeuGlyAlaSerV 605
|||||
354 ATTGAAGTCCCACTTAATGATTCAGAGATTCAGAGCTTGCTGCAGG 403
605 aLlyGlyAsnArgSerLysGluAsnHisAlaAspLeuGlyIlePheVal 621
|||||
404 TCAAAGGTAAACGGTCAAAAGAAACACGACGATTTGGGAATCTTTGTC 453
622 LysSerIleIleAsnGlyGlyAlaAlaSerLysAspGlyArgLeuArgVa 638
|||||
454 AAGTCATTATTAAATGAGAGACAGCATCTAAAGATGCAAGCCTCGGCT 503
638 lAsnAspGlnLeuIleAlaValAsnGlyLysSerLeuLeuGlyThrA 655
|||||
504 GAATGATCAACTGATAGCAGTAATGAGAAATCCCTGGGCAAGACAA 553
655 sngLAspAlaMetGluThrLeuArgArgSerMetSerThrLysGluAsn 671
|||||
554 ACCAAGATGCCATGGAACCTTAAGAAAGGTATGCTACTAAGGCAAT 603
672 LysArgGlyMetIleGlnLeuIleValAlaArgArgIleSerLysCysAs 688
|||||
604 AAACGAGATGATCCAGCTTATTTGTGCAGAGAGATAGCAAGTGCA 653
688 nGluLeuLysSerProGlySerProProGlyProGluLeuProIleGluT 705
654 TGAGGTGAAGTACCTGGGAGCCCTGGACCTGAGCTGAGCCCAATGAAA 703
705 hValLeuAspAspArgGluArgArgIleSerHisSerLeuThrSerGly 721
|||||
704 CAGCGTGGATGATAGAGAAACGAAAGATTTCCATTCCTCTACAGTGG 753
722 IleGluGlyLeuAspGluSerProSerArgAsnAlaAlaLeuSerArgI 738
|||||
754 ATTGAGGGGCTTGATGAATCGCCCAAGCAAAATGCGCCTCAGTAGAT 803
738 eMetGlyLysGlyLysTyrGlnLeuSerProThrValAsnMetProG 755
|||||
804 AATG.....GGTAAATACAGCTGTCCTCCCTACAGTAATATGCCCC 844
755 lAspAspThrValIleIleGluAspAspArgLeuProValLeuProPro 771
|||||
845 AAGATGACACCTGCTATTATAGAAAGATGACAGGTGCAAGTCTCTCCA 894
772 HisLeuSerAspGlnSerSerSerSerSerHisAspAspValGlyPheVa 788
|||||
895 CATCTCTGACACAGTCCTTCCAGCTCCCATGATGATGGGGTTGCT 944
788 lThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIleSerAspSerAla 805
|||||
945 GACGGCAGATGCTGTACTTGGGCCCAAGCTCTCAATCACTGATTCAGCCG 994
805 sPCysSerLeuSerProAspValAspProValLeuAlaPheGlnArgGlu 821
|||||
995 ACTGCTCTTGAGTCAGATGTTGATCCAGTCTTCTTTCAACAGAA 1044
822 GlyPheGlyArgGlnSerMetSerGluLysArgThrLysGlnPheSerAs 838
|||||
1045 GGATTTGAGCGTCAGAGTATGTCAGAAACACGACAAACCAATTTTCAGA 1094
838 palAserGlnLeuAspPheValLysThrArgLysSerLysSerMetAspL 855
```

|||||
1095 TGCAGCAATTGGATTTCCTTAAACACCAAAATCAAAAGCATGATT 1144
855 eucllyilealaaspgluThrThrLeuasnThrValaAspaspGlnLysAla 871
1145 TAGCTATAGCTGACGACGACTAACCTCAATACAGTGATGACCGAAGCA 1194
872 GlySerProSerArgAspValGlyProSerLeuGlyLeuLysLysSerSe 888
1195 GGTTCTCCACGACAGATGTGGGTCTCTCCCTGGGTCTGAAGAAGTCAAG 1244
888 rSerLeuGluSerLeuGlnThrAlaValaGluValThrLeuasnGlyAla 905
1245 CTGCTTGGAGAGTCTGCAGACCCAGTTCGCCGAGTCACTTGAATGGCG 1294
905 spIleProPheH1sArgProArgProArgIleIleArgGlyArgGlyLys 921
1295 ATATTCTCTTCCATCGTCCACGGCCGGGATTAATCAGAGCGCAGGGGATGC 1344
922 AsnGluSerPheArgAlaAlaIleAspLysSerTyrAspLysProAlaVala 938
1345 AATGAGAGCTTACAGAGCTGCCATCGCAAAATCTTATGATTAACCCCGGT 1394
938 IAspAspAspAspGluGlyMetGluThrLeuGluGluAspThrGluGlu 955
1395 AGATGATGATGATGAAGGCATGAGACCTTGGAAGAACACAGAACAAAA 1444
955 eSerArgSerGlyArgGluSerValSerThrAlaSerAspGlnProSer 971
1445 GTTCAAGATCAGGAGAGACTGTGTATCCACAGCCAGTGTACGCCCTTCC 1494
972 HisSerLeuGluArgGlnMetAsnGlyAsnGlnGluLysGlyAspLysTh 988
1495 CACTCTCTGAGAGACAAATGAATGGAACCAAGAGAAAGTGTATTAGAC 1544
988 rAspArgLysLysAspLysThrGlyLysGluLysLysLysAspArgAspL 1005
1545 TGATAGAAAAAGATAAACTGAAAAAGAAAGAAAGATAGAGATA 1594
1005 yAsGluLysAspLysMetLysAlaLysLysGlyMetLeuLysGlyLeuGly 1021
1595 AGGAGAAAGATTAATAATGAACCAAGAGAGATGCTGAAGGCGCTTGGA 1644
1022 AspMetPheArgPheGlyLysHisArgLysAspAspLys 1034
1645 GACATGTTCAGCTGCCAAACTGAAGCCGAGAGAGAGA 1683
seq_name: /SID1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA527320
seq_documentation_block:
ID AA527320 standard; cDNA; 823 BP.
XX
AC AA527320;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, seq ID 355.
XX
KM Neuroprotective; cytosolic; dermatological; immunosuppressive; tumour;
KM antiinflammatory; anti-HIV; antibacterial; antinflammatory; cancer;
KM immune system disorder; rheumatoid arthritis; inflammatory condition;
KM organ transplant rejection; infection; hepatitis C; blood disorder;
KM sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KM neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KM chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KM cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KM reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KM acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1.
XX

PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0198874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205315.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 21-SEP-2000; 2000US-0234714.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.

393 eValaInSerAlaGlyLeuHisThrValGlnArgAlaProArgLeuAsn 409
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 406 GTGTGAACAGTGCAGGGCTTCACACGGTGCAGAGACCCCGACGTGAAC 455
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 410 HisProProGluGlnIleAspSerHisSerArgLeuProHisSerAlaH1 426
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 456 CACCCGCTGAGCAGATGACTCTCATCTCAAGACTACCTCATAGCCGACA 505
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 426 sProSerGlyLysProProSerAlaProAlaSerAlaProGlnAsnValP 443
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 506 CCGCTGGGAAAAACACATCCGCTCCAGCTGGCACCCTCAAAATGTAT 555
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 443 heserThrThrValSerSerGlyTyrAsnThrLysLysIleGlyLysArg 459
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 556 TTAGTACGACTGTACGACAGTGTATATACACAAAAAAATAGCAAGAG 605
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 460 LeuAsnIleGlnIleuLysLysGlyThrGluGlyLeuGlyheserIleTh 476
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 606 CTTAATATCCAGCTTAAGAAAGGTACAGAAAGTTGGGATTCAGCATCAC 655
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 476 rSerArgAspValThrIleGlyLysSerAlaProIleTyrValLysAsn1 493
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 656 TTCCAGAGATGTAAACAATAGTGGCTCAGCTCCAATCTATGTAAAAACA 705
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 493 IeLeuProArgGlyAlaAlaIleGlnAspGlyArgLeuLysAlaGlyAsp 509
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 706 TTCTCCCGCGGGCGGCGCATTCAGAGATGCGCAGCTTAAGCAGAGAGAC 755
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 510 ArgLeuIleGlnValAsnGlyValAspLeuValGlyLysSerGlnGluG1 526
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 756 AGACTTATATGAGGTAAATGAGTTTATGTTAGTGGCAAAATCCCAAGAGA 805
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 526 uValValSerLeuLeu 531
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 806 AGTTGTTGCGCTGTG 821
 |||||||||||||||||||||||||||||||||||||||||||||||||||

seq_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH33594

seq_documentation_block:
 ID AAH33594 standard; cDNA; 823 BP.
 XX
 AC AAH33594:
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:650.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KM colorectal carcinoma; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPT: 2001-235357/24.
 DR P-PSDB: AAG74163.
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 1; Page 2675; 9803pp; English.

XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytosolic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SO Sequence 823 BP; 249 A; 183 C; 201 G; 188 T; 2 other;

alignment_scores:
 Quality: 1397.00 Length: 272
 Ratio: 5.174 Gaps: 0
 Percent Similarity: 99.265 Percent Identity: 99.265

alignment_block:
 US-09-757-781-2 x AAH33594 ..

Align seg 1/1 to: AAH33594 from: 1 to: 823

260 LeuGlnHisIleProAsnPheserLeuAspAspMetValLysLeuValG1 276
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 6 TTGGAGCATATACCCCAACTTCTCTGTGATGATATGTAAGCTCGTAGA 55
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 276 uValProAsnAspGlyLysProLeuGlyIleHisValValProPhesera 293
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 56 AGTCCCAACAGATGGAGGGCTCTGGGAATCCATGTAGTGGCTTTCAGTG 105
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 293 IaArgGlyGlyArgThrLeuGlyLeuValLysArgLeuGlyLysGly 309
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 106 CTCGAGGCGGCGAGAACCCCTGGGTTATTAGTAAACCATTTGCAGAAAGCT 155
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 310 GlyLysAlaGlnHisGluAsnLeuPheArgGlnAsnAspCysIleValAr 326
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 156 GGTAAAGCTGAACTGAATAATCTTTTCGTGAGATATATATGTCATGTCAG 205
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 326 gIleAsnAspGlyAspLeuArgAsnArgArgPheGlnGlnAlaGlnHisM 343
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 206 GATTAAATGATGGCAGCTTCGAATAGAGATTGTGAACACAGACACACTA 255
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 343 etPheArgGlnAlaMetArgThrProIleIleTrpPheHisValValPro 359
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 256 TGTTCGCCAAGCCATGCGTACACCATCATTTGTTCCATGTGTTGCT 305
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 360 AlaAlaAsnLysGluGlnIleuSerGlnLeuSerGlnIleuLysAsnAs 376
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 306 GCAGCAATATAAGAGCAGATGAACTATCCCAAGTGAAGAACAA 355
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 376 nTyrTyrSerSerArgPheSerProAspSerGlyTyrIleAspAsnArgS 393
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 356 TTACTATTCAAGCCGTTTATAGCCCTGACAGCCATATATTATACAAACGA 405
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 393 eValaInSerAlaGlyLeuHisThrValGlnArgAlaProArgLeuAsn 409
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 406 GTGTGAACAGTGCAGGGCTTCACACGGTGCAGAGACCCCGACGTGAAC 455
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 410 HisProProGluGlnIleAspSerHisSerArgLeuProHisSerAlaH1 426
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 456 CACCCGCTGAGCAGATGACTCTCATCTCAAGACTACCTCATAGCCGACA 505
 |||||||||||||||||||||||||||||||||||||||||||||||||||

CC can be used as targets in developing anti-cancer agents such as
 CC chemotherapeutic compounds. The markers can also be used as targets in
 CC developing treatments for cancer, particularly those cancers which
 CC display resistance to agents and exhibit expression of the markers. The
 CC anticancer agents developed by the novel method can be used to treat
 CC cancer. Probes based on the markers can be used to detect transcripts or
 CC genomic sequences corresponding to the markers, in the identification of
 CC cells or tissues which mis-express the protein. Cancers which may
 CC be targeted include carcinoma (e.g. squamous cell carcinoma),
 CC sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
 CC lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
 CC tumours (e.g. glioma). The present sequence is one of the 1046
 CC novel cancer cell markers.

XX
 SQ Sequence 2100 BP; 649 A; 460 C; 544 G; 439 T; 8 other;

alignment_scores:
 Quality: 1187.50 Length: 849
 Ratio: 2.385 Gaps: 25
 Percent Similarity: 58.657 Percent Identity: 36.867

alignment_block:
 US-09-757-781-2 x AAS60861 ..

Align seg 1/1 to: AAS60861 from: 1 to: 2100

```

seq_name: /SIDSL/jcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AAS60861
seq_documentation_block:
ID   AAS60861 standard; cDNA; 2100 BP.
XX
XX   AAS60861;
AC
XX
XX   29-JAN-2002 (first entry)
DE
XX   Human cancer agent-resistance marker #520.
XX
XX   Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
XX   squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
XX   lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
XX   Hodgkin's disease; glioma; ss.
XX
XX   Homo sapiens.
OS
XX
XX   WO200179556-A2.
XX
XX   25-OCT-2001.
XX
XX   13-APR-2001: 2001WO-US12132.
XX
XX   14-APR-2000: 2000US-197538P.
XX
XX   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX   Lillie J, Brown JL, Bolt A, Van Huuffel C;
XX
XX   WPI; 2001-602933/68.
XX
XX   Novel nucleic acid, used as a marker to determine the effectiveness of
XX   using TAXOL to treat cancer cell growth in individuals -
XX
XX   Claim 1; Page 370; 527pp; English.
XX
XX   The invention relates to 1046 novel nucleic acids which are used as
XX   markers for determining the sensitivity of a cancer cell to the
XX   anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
XX   they are shown to express one of the 242 sensitivity markers or the
XX   cells are shown not to express one of the 804 resistance markers.
XX   The methods can be used to determine the effectiveness of TAXOL
XX   in the treatment of cancer cell growth in an individual. The markers
  
```

```

223 SerAlaSerHisProMetValGlyIleTrpLeuGluysGingInAs 239
    |||||||
49 AGCGCTTCACACCTCG.....TGCTTAGAGTGTGAACCGGATG 88
    |||
239 pGluAspGlyThrGluGluAspAsnSerArgValGluProValGlyHisA 256
    :: |||
89 TTCAGAGGTACAGCAGAACTTGGAAGACAGA...GAACTTTGATGGTGC 135
    |||
256 lAspThrGlyLeuGluHisIleProAsnPro.....SerLeuAspAsp 270
    ::||| |||
136 TCACAGACAGAACTACTAACTTCGCCAAGACACTAGGACACATGTGAT 185
    |||
271 MetValIleValGluValProAsnAspGlyGlyProLeuGlyIleHis 287
    |||
186 AGCAGACAGAAAGAGTGGAGATTTCTGGGAAAGAGGCCCATTTGGCAAT 235
    |||
287 sValValPro...PheSerAlaArgGlyLysArgThrLeuGlyLeuVal 303
    |||
236 TGTAGTGGCCCTTTCTTCACTGTGAGTGAAGATCTTAGGACTCTTCA 285
    |||
303 AlIysArgLeuGluysGlyGlyLysAlaGluHisGluAsnLeuPheArg 319
    ::|||
286 TCAGAGGCAATTGACAGACACACAGAGTCCAGCGGAGGCAATTTTCAC 335
    |||
320 GluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnArgAr 336
    |||
336 GAAATGATGATGTATGTAAATAATCAACAATGTGAGATCTGTAGACAAAC 385
    |||
336 grheGluGluAlaGluHisMerPheArgIleAlaMetArgThrProIle 353
    |||
386 CTTTGCTCAGGCTCAAGATGTCTCCGACAGGCAATGAATATCCCAAGTG 435
    |||
353 lerrPheHisValValProAlaAlaAsnLysGluGluIntyrGluGluLeu 369
    :: |||
436 TGCTCCGTCAGCGTCTCCACAAACCGTGACAGATATGAAAAG... 482
    |||
370 SerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerProAspse 386
    |||
483 TCAGTCATTCGCTCTTAACATTTTGTAAATGAATGATGGCTTTTGAA 532
    |||
386 rGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHisThrValG 403
    :: |||
533 AACCAAGTGGCCGCTCCATCATGAAATGGGACATAACACA.... 578
    |||
403 lnaGAlaProArgLeuAsnHisProGluGluHisAspSerHisSer 419
    |||
  
```

```

579 .....GCAATCTCACAGGACCGATAGTCTTCAACAGATCATCACT 623
420 ArgLeuProHisSerAlaHisProSer.....GlyLysProP 432
    ||| ||| |||
624 TCCCTGCACAAACAAAGAGTCCCGGAGTACAAAGGCTGGAGAGAAACC 673
432 oSerAlaProAlaSerAlaProGlnAsnValPheSerThrThrValSerS 449
    ||| ||| |||
674 ATCTCTCCCTCACTCTCCCT.....CTCA 639
449 erGlyThrAsnThrLysLysLLeGlyLysArgLeuAsnLLeGlnLeuLys 465
    ||| ||| |||
700 TGGGATTTGGACCAATATAAAATGCAAAATAAATTAGATTGACCTTAAG 749
466 LysGlyThrGlnGlyLeuGlyPheSerLLeThrSerArgPValThrII 482
    ||| ||| |||
750 AAAGGCCCTTAAGAGACTGTGTTCACGTGTACCGAGACACTTCCAT 799
482 eGlyGlySerAlaProLLeTyrValLysAsnLLeLeuProArgGlyAlaA 499
    ||| ||| |||
800 ACATGTGTCCTCCGTCCTATTTTGTAAAAAACATTTTACCAAGGAGCAG 849
499 LaLLeGlnAspGlyArgLeuLysAlaGlyAspArgLeuLLeGlyValAsn 515
    ||| ||| |||
850 CTATATTAAGATGGCCGCTACAAATCAGGGGACAGAAATTTTGAGGTAAAT 899
516 GlyValAspLeuValGlyLysSerGlnGlyValValAspSerLeuLeuA 532
    ||| ||| |||
900 GGGAGAGATGTCAACCGGACGACCCAGAGAGACTTGTGGCCATTCCTCAG 949
532 gSerThrLysMetGlnGlyThrValSerLeuLeuValPheArgGlnGlyA 549
    ||| ||| |||
950 GACACCCAAACAGGGGAGACACATCCCTCATATGCCCCGCCAAGAG 999
549 sPAlaPheHisProArgGlnLeuAsnAlaGluProSerGlnMetGlnLe 565
    ||| ||| |||
1000 GACATTTTCTGCCCGAGAGTTG..... 1022
566 ProLysGlnThrLysAlaGlnAspGlnAspLLeValLeuThrProAspG 582
1022 ..... 1022
582 yThrArgGluPheLeuThrPheGluValProLeuAsnAspSerGlySerA 599
1022 ..... 1022
599 LaGlyLeuGlyValSerValLysGlyAsnArgSerLysGlnAsnHisAla 615
1022 ..... 1022
616 AspLeuGlyLLePheValLysSerLLeLeaGlyLysAlaAlaSerLys 632
1022 ..... 1022
632 sAspGlyArgLeuArgValAsnAspGlnLeuLLeAlaValAsnGlyLys 649
    ||| ||| |||
1023 .GATGGTCTCTCGAATGATGATCAGCTGATTCAGTTAATGGGGAAT 1071
649 erLeuLeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSer 665
    ||| ||| |||
1072 CTCTTTTGGAAAGTCCACACAGACCTATGGAAACACTTATGGCGGTCA 1121
666 MetSerThrGlnGlyAsnLysArgGlyMetLLeGlnLeuLysAlaAlaA 682
    ||| ||| |||
1122 ATGTCATCGAGGAAACATCCGAGGATGATCCAGTTGGTGATTCGAC 1171
682 gaGlyLLeSerLysCysAsnGluLeuLysSerProGlySerProGlyP 699
    ||| ||| |||
1172 GAGG.....C 1176
699 roGluLeuProLLeGluThrAlaLeuAspAspArgGluArgGlyLLeSer 715
    ||| ||| |||
1177 CAGAGAGACCAATG..... 1190

```

```

716 HisSerLeuTyrSerGlyLLeGlyLysAspGlnSerProSerArgAs 732
    ||| ||| |||
1191 .....GAGATCTCTGAGAGTG 1207
732 nAlaAlaLeuSerArgLLeMetGlyLysSerGlyLysTyrGlnLeuSerP 749
    ||| ||| |||
1208 TGGGCAATTTTCCAAAGCCATGCTTGTGAGAAC.....TGCCAAATGCTG 1251
749 roThrValAsnMetProGlnAspAspThrValLLeLLeGluAspAspArg 765
    ||| ||| |||
1252 TAAACACCTCTAGGCGAAATGATATATAGTATC..... 1283
766 LeuProValLeuProPronHisLeuSerAspGlnSerSerSerSerSerH 782
    ||| ||| |||
1284 .....CTGCATCCACTTGGCACTTGCACATCCACA 1312
782 sAsp..AspValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysA 798
    ||| ||| |||
1313 AGACMAACAGAAAGTCTATGTGTGCCCAATGACGGA...TGGGCCGAGA 1359
798 LaAlaLLeSerAspSer..... 803
1360 GTCAAGTTCACCTTCTCCAAACACACATTCGTCTGGGATTGGGCTC 1409
804 AlaAspCysSerLeuSerProAspValAspProValLeuAlaPheGlnA 820
    ||| ||| |||
1410 GAAGATTTAACAGCCACACTCTGGGGTGATTCACACAGATATTTTCCAGA 1459
820 gGluGlyPheGlyArgGlnSerMetSerGlnLysArgThrLysGlnPheS 837
    ||| ||| |||
1460 TCACACACATCAACTTCAGATCTGTGACCCGCGCAG..... 1496
837 erAspAlaSerGlnLeuAspPheValLysThrArgLysSerLysSerMet 853
1497 .....CAGCCTGAATCAATTAATTTGAAAGCCTCGAAGAGCATG 1535
854 AspLeuGlyLLeAlaAspGlnThrLysLeuAsnThrValAspAspGlnLy 870
    ||| ||| |||
1536 GACCTT...GTGCCAGATGAAGAGAGAGTTCATCATTCGTGGACGACAAA 1582
870 sAlaGlySerProSerArgAspValGlyProSerLeuGlyLeuLysLys 887
    ||| ||| |||
1583 ATCGGAATCTCCAAAGAAAGATTTTGTGCCAACTGTGGTTGAAAAAGT 1632
887 erSerSerLeuGlnSerLeuGlnThrAlaValAlaGluValThrLeuAsn 903
    ||| ||| |||
1633 CCAGCTCTTGGAGAGTCTGCAGACTGCAGTGCAGAGTCAGAAAGAT 1682
904 GlyAspLLeProPheHisArgProArgProArgLLeLLeArgGlyArgG 920
    ||| ||| |||
1683 ...GACCTTCTCTTCAACAGGCCCGCGCGCACATGTGTCAGGCCGAGG 1729
920 yCysAsnGlnSerPheArgAlaAlaLLeAspLysSerTyrAspLysProA 937
    ||| ||| |||
1730 CTGCAAATGAGAGCTTTAGAGCAGCCATTCACMAATCTCACATGAGACT 1778
937 LaValAspAspAspGlnGlyMetGluThrLeuGlnGluAsp..... 951
1779 .....CAGAAATAGAAAGCTGACGAGTCTG 1802
952 ThrGlnLysSerArgSerGlyArgGlnLysSerValSerThrAlaSerAs 968
    ||| ||| |||
1803 TCTGATTAAGACTCTCACTGTGCCAAGGAGACTCTGAATGTGAGTCTGC 1852
968 pGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGlyLysG 985
    ||| ||| |||
1853 CCTCAGGGGGAATTCGAGCTAGAGACATGGAATAAATAAACCCAGAAAG 1902
985 LysPlyLysThr.....AspArgLysLysAspLys.....Thr 995
    ||| ||| |||
1903 TCAAAAAAAGCAGAAAGAGAGAGAAAGAAAAAGAAAGGCGCAATTGAAA 1952

```

```
996 GilysglyLysLysAspArgAspLysGlyLysAspLysMetLysAl 1012
|||||
1993 GTCAAGGAGAAAAGCGCAAGAGATGAAGATCCAGAAAGCAAAAT 2002
|||||
1012 alysLysGlyMetLysGlyLysAspMetPheArgPheGlyLysH 1029
|||||
2003 AAGAAG.....AAGGCTCGCGCCATGCTGAGATTGGAAAGA 2043
1029 lAsArgLysAsp.....AspLysIleGlyLysThrGlyLysIleLys 1042
|||||
2044 ACAAAGAGATGAAGGTGGAAGCGCTGACAGAAAGTACTCTGAAA 2090
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NM2001A.DAT:AA560956
seq_documentation_block:
ID   AA560956 standard; cDNA: 2100 BP.
XX
AC   AA560956;
XX
DT   29-JAN-2002 (first entry)
XX
DE   Human cancer agent-resistance marker #615.
XX
KW   Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW   squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW   lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW   Hodgkin's disease; glioma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200179556-A2.
XX
PD   25-OCT-2001.
XX
PE   13-APR-2001; 2001WO-US12132.
XX
PR   14-APR-2000; 2000US-197538P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J, Brown JL, Bolt A, Van Hufel C;
XX
WP1: 2001-602933/68.
XX
PT   Novel nucleic acid, used as a marker to determine the effectiveness of
XX   using TAXOL to treat cancer cell growth in individuals -
XX
PS   Claim 1; Page 458-459; 527pp; English.
XX
CC   The invention relates to 1046 novel nucleic acids which are used as
CC   markers for determining the sensitivity of a cancer cell to the
CC   anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
CC   they are shown to express one of the 242 sensitivity markers or the
CC   cells are shown not to express one of the 804 resistance markers.
CC   The methods can be used to determine the effectiveness of TAXOL
CC   in the treatment of cancer cell growth in an individual. The markers
CC   can be used as targets in developing anti-cancer agents such as
CC   chemotherapeutic compounds. The markers can also be used as targets in
CC   developing treatments for cancer, particularly those cancers which
CC   display resistance to agents and exhibit expression of the markers. The
CC   anticancer agents developed by the novel method can be used to treat
CC   cancer. Probes based on the markers can be used to detect transcripts or
CC   genomic sequences corresponding to the markers, in the identification of
CC   cells or tissues which mis-express the protein. Cancers which may
CC   be targeted include carcinoma (e.g. squamous cell carcinoma),
CC   sarcoma (e.g. fibrosarcoma) leukaemia (e.g. squamous cell leukaemia),
CC   lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
CC   tumours (e.g. glioma). The present sequence is one of the 1046
XX   novel cancer cell markers.
XX
SQ   Sequence 2100 BP; 649 A; 460 C; 544 G; 439 T; 8 other;
```

```
alignment_scores:
Quality: 1187.50      Length: 849
Ratio: 2.385          Gaps: 25
Percent Similarity: 58.657      Percent Identity: 36.867
```

```
alignment_block:
US-09-757-781-2 x AA560956 ..
```

```
Align seg 1/1 to: AA560956 from: 1 to: 2100
```

```
223 SerAlaSerHisPheMetValGlyLysThrPheLysGlyGlnGlnLys 239
|||||
49 AGCGGCTTCACACCTGG.....TGGTAGAGTCTGAAACCGGATG 88
239 pGluAspGlyThrGlnGluAspAsnSerArgValAlaGluProValGlyHisA 256
|||||
89 TTCACAGGCTCACGCGCAAGACTTGGAAAGACAGA..GAAGTTTGATGGTG 135
256 lAspThrGlyLeuGlnHisIleProAsnPhe.....SerLeuAspAsp 270
|||||
136 TACAGACAGAACTACTTACTTCGCCAAGAACTAAGGACATGTAGAGTAT 185
271 MetValLysLeuValGluValProAsnAspGlyLysProLeuGlyLysI 287
|||||
186 ATGACAAAGAACAGTGTCTGTGGGAAAGAGGCCCATTTGGAAATGAC 235
287 sValValPro...PheSerAlaArgGlyLysArgThrGlyLysLeuVal 303
|||||
236 TGTAGTGGCCCTTTTTCATCTCTGAGTGGAAAGATCTAGACCTTCTCA 285
303 alLysArgLeuGlyLysGlyLysAlaGlnHisGlnAsnLeuPheArg 319
|||||
286 TCCGAGGACATGTGACACACACAGAGTCCACGCGGAGGACTATTTTCAC 335
320 GluAsnAspCysIleValArgIleAsnAspGlyAspLeuArgAsnArg 336
|||||
336 GAAATGAATGTATGTGTAAATCAACAATGTGATGTGTGACAAAC 385
336 pPheGlnGlnAlaGlnHisMetPheArgGlnAlaMetArgThrProIle 353
|||||
386 CTTCGCTCAGGCTCAAGATGTCTCTCCGCGCAAGATGAATGTCCACAGTG 435
353 lEtrPheHisValValProAlaAlaAsnLysGlnIleThrGlnLys 369
|||||
436 TGTCTCTCCACGTCGTCCTCCACAAACCGTCAACAGATGATAAAG... 482
370 SerGlnSerGlnLysAsnAsnLysThrSerArgPheSerProAsp 386
|||||
483 TCGATCATTTGGCTCTTAACTTTTGGTAATGATGAGCGCTTTGAA 532
386 rGlnIleAspAsnArgSerValAsnSerAlaGlyLeuHisThrValG 403
|||||
533 AACCAAAAGTCCGCGCTCTGTCCATGGAATAATCGGACATAAGACA.... 578
403 lAsArgAlaProArgLeuAsnHisProProGlnIleAspSerHisSer 419
|||||
579 .....GCAATCTCACAGGACGAGATAGTCTGAAACAGATGATCAGCT 623
420 ArgLeuProHisSerAlaHisProser.....GlyLysProPr 432
|||||
624 TCCCTCCAAACAAACAAAGTCCCGACAGTACCAAGCTGGAGGAAAC 673
432 oSerAlaProAlaSerAlaProGlnAsnValPheSerThrThrValSer 449
|||||
674 ATGCTCTCCCTCAGCTCGCT.....CTCA 699
449 eArgLysAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLys 465
|||||
700 TGGGATTTGGCAGCAATTAATAAATGCAAAAGAAATTAAGATTGACCTAAAG 749
466 LysGlyThrGlnGlyLeuGlyPheSerIleThrSerArgAspValThrI 482
|||||
750 AAAGGCCCTGAAGACTTGGTTTCACTGTGTTACCAAGACATCTTCCAT 799
```

```

482 eGIyGYSerIaProIleTyValLysAsnIleLeuProArgIyAlaA 499
    ||| :::::::::::|||||
800 ACATGGTCCCGCCATTTTGTAAAAACATTTTACCAGAGAGAG 849
499 IalIeIlnspsGIyArgLeuYsAlaGIyAspArgLeuIleGIuValAsn 515
    |||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
850 CTAATAAATGAGCCCGCTACAAATAGGGGAGCAATTTTGGAGCTAAT 899
516 GIYValAspLeuValGIySserGIuGIuValValSerLeuLeuAr 532
    ||| ||| :::::::::::||||| ||||| ||||| ||||| |||||
900 GSGAGAGATGTCACCGAGCAACCCAGAGAGCTGTGGCCATGCTCAG 949
532 gSerThrLyMetGIuGIYThrValSerLeuLeuValPheArgGIuGIuA 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
950 GACACCAAGCAGGGGAGACACGATCGCTGTCTATCCCGCCAGAGAG 999
549 sPaIaPheHisProArgGIuLeuAsnAlaGIuProSerGIuMetGIuIle 565
    ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1000 GACATTTTCTGCCCGAGAGGTG..... 10222
566 ProLYSGluThrLysAlaGIuAspGIuAspIleValLeuThrProAspGI 582
1022 ..... 10222
582 yThrArgGIuPheLeuThrPheGIuValProLeuAsnSserGIySera 599
1022 ..... 10222
599 IaGIyLeuGIyAlServalYsGIyAsnArgSerLyGIuAsnHisAla 615
1022 ..... 10222
616 AspLeuGIyIlePheValYsSerIleIleAsnGIyGIuAlaIaSerLy 632
1022 ..... 10222
632 sAspGIyArgLeuArgValAsnAspGIuLeuIleAlValAsnGIyGus 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1023 .GATGTCGTCTCGCAATGAATAGACCAGCTGATTTCACTTAATGGGAAAT 10711
649 eRIeuleuGIyLysThrAsnGIuAspAlaMetGIuThrLeuArgYsSer 665
1072 CICTTWTGGGAAAGTCCACACGAGACGTAATGGAACACTTAGGGGCA 11211
666 MetSerThrGIuGIyAsnArgGIuMetIleGIuLeuIleValAlaAr 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1122 AGTCCATGAGGAGAAACATCCGAGGAGATGCCGTTGGATTCAG 11711
682 gArgIleSerLysCysAsnGIuLeuYsSerProGIySerProProGIyP 699
1172 GAGG.....C 1176
699 roGIuLeuProIleGIuThrAlaLeuAspAspArgGIuArgIleSer 715
1177 CAGAGAGACCAATG..... 1190
716 HisSerLeuTySserGIyIleGIuGIyLeuAspGIuSserProSerArgAs 732
1191 .....GAGGATCTCTGCAGAGCTG 1207
732 nAlaIaIaIeuserArgIleMetGIyGIuSserGIyLysTyrgIuLeuSerP 749
    ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
1208 TGGGGCATTTTCCAGACCATGCTTGAGAAC.....TGTCAAAATGCTG 12511
749 roThrAlaIaIaIeuserProGIuAspAspThrValIleIleGIuAspArg 765
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
1252 TAACCACTCTTAGCGCAAAATGATAATGATC..... 12833
766 LeuProValIeuproProHisIeuserAspGIuSserSerSerSerHI 782
    ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1284 .....CTGATGCCACTTGGCACTTGAGAGGCCCA 1312

```

[illegible]

271 MetValIysLeuValGluValProAsnAspGlyGlyProLeuGlyTleH 287

180 ATGACAAAGACAGTGGAGATTCTGGGGAAGAGGCCCATTTGGATACA 235

207 svaLVLPfo...pheseraLaargGlyGlyargThrLeuGlyLeuLeuV 303

230 1GAGGCGCCCTCTTTCATCTCTGAGTGGAGGATCTTAGACTCTCA 285

200 ACCGAGGCACTGAAGACACACAGCAGGTCACAGCGGAGGACATTTCAC 335

336
 ...cysllvalargileasnaspglyasp leuargasnargar
 |||||::|||...
 ...cysllvalargileasnaspglyasp leuargasnargar 336

... CAGGATGTCCTCCGCGCAGGCATGAATCTCCAAGTG 435

...GCAAGAAATTAAGATTGACCTTAAG 749

AAAGGCCCTGAGGACTTGGTTCACTGTGGTTACCGAGAGACTCTTCAT 799

2 EGIYGLYSERALAPROIETyrValLysAsnIleLeuPROARGLVA]aa 499

0 ACATGGTCCCGGTCCTTTTGTAAAAACATTTTACCAAGGAGCAG 849

2 gSerThrLysmetGluGlyThrValSerLeuLeuValIleAspGluVal

GAGCACCAGCAGGGGAGACAGCATGGTATTCACCGCGG

PF 13-APR-2001; 2001MO-US12132.
XX
PR 14-APR-2000; 2000US-197538P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Little J, Brown JL, Bolt A, Van Huffel C;
XX WPI; 2001-602933/68.
XX
PT Novel nucleic acid, used as a marker to determine the effectiveness of
PS using TAXOL to treat cancer cell growth in individuals -
XX
XX Claim 1; Page 498; 527pp; English.

CC The invention relates to 1046 novel nucleic acids which are used as
CC markers for determining the sensitivity of a cancer cell to the
CC anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
CC they are shown to express one of the 242 sensitivity markers or the
CC cells are shown not to express one of the 804 resistance markers.
CC The methods can be used to express one of the 804 resistance markers.
CC In the treatment of cancer cell growth in an individual. The markers
CC can be used as targets in developing anti-cancer agents such as
CC chemotherapeutic compounds. The markers can also be used as targets in
CC developing treatments for cancer, particularly those cancers which
CC display resistance to agents and exhibit expression of the markers in
CC anticancer agents developed by the novel method can be used to treat
CC cancer. Probes based on the markers can be used to detect transcripts or
CC cells or tissues which mis-express the markers. In the identification of
CC be targeted include carcinoma (e.g. squamous cell carcinoma),
CC lymphoma (e.g. fibrosarcoma), leukemia (e.g. lymphocytic leukemia),
CC tumours (e.g. glioma), reticulum cell sarcoma, Hodgkin's disease and
CC novel cancer cell markers.
XX
SO Sequence 2100 BP; 649 A; 460 C; 544 G; 439 T; 8 other;

alignment_scores:
Quality: 1187.50
Ratio: 2.385 Length: 849
Percent Similarity: 58.657 Gaps: 25
Percent Identity: 36.867

alignment_block:
US-09-757-781-2 x AAS61007

Align seg 1/1 to: AAS61007 from: 1 to: 2100

223 SerAlaSerHisProMetValGlyLysTrpLeuGluLysGlnGluGlnAs 239
|||
49 AGCGCTTCACACCCCTGG.....TGGTTAGAGTCTGAAACCGGATG 88
239 pGluAspGlyThrGluGluAspAsnSerArgValGluProValGlnHisA 256
89 TTTCCAGGGTCCACCAAGAACTGGAGACAGA...GAAAGTTTGAATGGTG 135
256 LaaSPHrGlyLeuGluHisIleProAsnPro.....SerLeuAspAsp 270
136 TACAGACAACTACTACTTCGCCAAGAACTAAGACACATTGAGTGTAT 185
271 MetValLysLeuValGluValProAsnAspGlyLysProLeuGlyLeuH 287
186 ATGACAAAGACAGTGGAGATTCTGGGAGAGAGGCCCATTTGGGAATACA 235
287 sValValPro...PheSerAlaArgGlyLysArgThrLeuGlyLeuLeuV 303
236 TGTAGTGCCCTTTTCATCTCTGAGTGGAGGATTTCTAGAGCTCTTCA 285
303 aLysArgLeuGluLysGlyLysAlaGluHisGluAsnLeuPheArg 319
286 TCCGAGCATTTGAACACAGCAGGTCACAGCGGAGAGACTATTTCAC 335

320 GluAsnAspCysIleValArgIleAsnAspCysLysLeuArgAsnArgAr 336
|||
336 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 385
336 pPheGluGlnIleAsnHisMetPheArgIleValMetArgThrProIle 353
386 GTTCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAG 435
353 IetPheHisValValProAlaAlaAsnLysGlnGlyLeuGluLeu 435
436 TGTCTCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAGGCTCAG 369
370 SerGlnSerGluLysAsnAsnTrpTrpSerSerArgPheSerProAspSe 386
483 TCAGTCATGCTGCTCTTAACTTTTGTATATGATGCTGCTTGA 532
386 rGlnTrpIleAspAsnArgSerValAsnSerAlaGlyLeuHisThValG 403
533 AACCAAGTCCGCTCCTGTCATGCAATGCAATGCAATGCAATGCAATGCA 578
403 LnaArgAlaProArgLeuAsnHisProProGluGlnIleAspSerHisSer 419
579GCAATCTCAGAGAACGATGCTGCAATGCAATGCAATGCAATGCA 623
420 ArgLeuProHisSerAlaHisProSer.....GlyLysProPr 432
624 TCCCTGCAACAAACAGAGTCCCGAGTACCAAGGCTGGAGGAAAC 673
432 oSerAlaProAlaSerAlaProGlnAsnValPheSerThrValSers 449
674 ATCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 449
449 eArgLysAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLys 465
700 TGGGATTTGGGAGCAATATAAATGCAAGAAATTAATTAATTAATTAATTA 749
466 LysGlyThrGluGlyLeuGlyPheSerIleThrSerArgPheValThrI 482
750 AAGGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 799
482 eGlyLysSerAlaProIleTrpValLysAsnIleLeuProArgIleAla 499
800 ACATGCTCCGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 849
499 IalIleGlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsn 515
850 CTATATTAAGATGGCCCTTCAATCAGGAGGAGGAGGAGGAGGAGGAGGAG 899
516 GlyValAspLeuValGlyLysSerGlnGluValValSerLeuLeuAr 532
900 GGGAGAGATGTCACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 949
532 gSerThrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluA 549
950 GAGCACCAG 999
549 sPAlaPheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIle 565
1000 GACATTTCTGCTCCGAGAGAGTGG..... 1022
566 ProLysGluThrLysAlaGluAspGluAspIleValLeuThrProAspG 582
1022 1022
582 YThrArgGluPheLeuThrPheGluValProLeuAsnAspSerGlySera 599
1022 1022
599 IagLysLeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAla 615
1022 1022
616 AspLeuGlyIlePheValLysSerIleLeuAsnGlyLysAlaAlaSerLys 632

```

1022 ..... 1022
632 sAspGIyArGLeuArGLValAsnAspGlnLeuIleAlaValAsnGlyGlu 649
||||| ..... |||||
1023 .GATGTCGCTCTCGAATGATGATGACCAAGCTGATTCAGATTAAATGGGAAAT 1071
649 erLeuLeuGLyLysThrAsnGlnAspAlaMetGluThrLeuArGLysSer 665
||||| ..... |||||
1072 CTCTTTGGGAAAGTCCAAACGACGAGCTATGSAACACTTACGGCGTCA 1121
666 MetSerThrGluGlyAsnLysArGLysMetIleGlnLeuIleValAlaAr 682
||||| ..... |||||
1122 ATGTCATGAGGAGGAAACATCCAGAGGATGATCAGTTGATTCGAG 1171
682 gaRGLysSerLysCysAsnGlnLeuLysSerProGlySerProGlyP 699
||||| ..... C 1176
1172 GAGG.....
699 roGluLeuProIleGluThrAlaLeuAspAraGLuArGLysSer 715
||||| .....
1177 CAGAGGACCAATG..... 1190
716 HisSerLeuTyrSerGlyIleGluGlyLeuAspGlnSerProSerArGL 732
||||| ..... GAGGATCTCCAGAGTG 1207
1191 .....
732 naLaAlaLeuSerArGLMetGlyGlnSerGlyLysTyrGlnLeuSerP 749
||||| ..... |||||
1208 TGGGGCATTTTCCAGCCATGCTTGAGAAC.....TGTCAAATGCTG 1251
749 roThrValAsnMetProGlnAspAspThrValIleIleGluAspAraGL 765
||||| .....
1252 TAACCACTCTGAGCGAAATGATATAGTATC..... 1283
766 LeuProValLeuProProHisLeuSerAspGlnSerSerSerSerH 782
||||| .....
1284 .....CTGCATCCACTTGGCCTTGCAGTCCACA 1312
782 sAsp...AspValGlyPheValThrAlaAspAlaGlyThrTrrAlaLysA 798
||||| ..... |||||
1313 AGCAAAACAGAAAGGTCTATTGCTGCCCAATGACGGA..TGGGCCGAGA 1359
798 laAlaIleSerAspSer..... 803
|||
1360 GTGAAGTTCCACCTCTCCAAACACACATTCCTGCTGGATGGGCCTC 1409
804 AlaAspCysSerLeuSerProAspValAspProValLeuAlaPheGlnAr 820
||||| ..... |||||
1410 GAAGATTACGCCACAGCTCTGGGGTGATTCAGACATATATTTTCCAGA 1459
820 ggGluGlyPheGlyArGLInserMetSerGluLysArGLThrLysGlnPheS 837
||||| ..... |||||
1460 TCAGCACATCACTTCAGATCTGTGACACGGCCAGG..... 1496
837 erAspAlaSerGlnLeuAspPheValLysThrArGLysSerLysSerMet 853
||||| ..... |||||
1497 .....CAGCTGATCAATTAAATTGAAGCCCTCGAAGACGATG 1535
854 AspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAspAspGlnLy 870
||||| ..... |||||
1536 GACCTT...GTGCCAGATGAAACAGAGTTCATTCATTTGCTGACAAAA 1582
870 saLaGLysProSerArGLAspValGlyProSerLeuGlyLeuLysLysS 887
||||| ..... |||||
1583 ATCGGAATCTCCAAAGAAATTTTGTCCAACTCTGGTTTGAAGAAAGT 1632
887 erSerSerLeuGlnSerLeuGlnThrAlaValAlaGluValThrLeuAsn 903
||||| ..... |||||
1633 CCAGCTCTCTGGAGAGCTGCACAGCTGAGCCGAGGAGAGAGAAAT 1682
904 GLyAspIleProPheHisArGLProArGLProArGLIleIleArGLyArGL 920
||||| ..... |||||

```

```

1683 ...GACCTTCTTTTCACAGGCCCCGGCCGACATGTTGAGGCGCAGG 1729
920 yCysAsnGlnSerPheArGLAlaAlaIleAspLysSerTyrAspLysProA 937
||||| ..... |||||
1730 CTGCAATGACAGCTTTGAGCAGCAGCATTCGCAAAATCCTACAGTAGACT. 1778
937 laValAspAspAspAspGluGlyMetGluThrLeuGlnGluAsp..... 951
||||| ..... |||||
1779 .....GAGCAAAATAGAAAGCTGACAGGCTG 1802
952 ThrGlnGlnSerSerArGLSerGlyArGLSerValSerThrAlaSerAs 968
||||| ..... |||||
1803 TCTGATAGAGCTCTCACTGTGGCAAGAGCTCTGATTTGAGTCTGC 1852
968 gGlnProSerHisSerLeuGlnArGLInMetAsnGlyAsnGlnGluLysG 985
||||| ..... |||||
1853 CCCTCAGGCGAATTCGAGAGCTAGAGACATGCAAAATAAAGCCAGGAAG 1902
985 LyAspLysThr.....AspArGLysLysAspLys.....Thr 995
||||| ..... |||||
1903 TCAAAAAAAGCAAGAGAGAGAGCAAAAGCAAAAGCGCAAAATGAAA 1952
996 GlyLysGluLysLysLysAspAraGLysGluLysAspLysMetLysAl 1012
||||| ..... |||||
1953 GTCAAGAGAGAAAGCCGAAAGAGAGAAATGATCCAGAAAGCAAAAT 2002
1012 aLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArGLPheGlyLysH 1029
||||| ..... |||||
2003 AAAGCAAG.....AAGGCTTCGGCGCCCATCTGATGATTTGGAAGA 2043
1029 tsArGLysAsp.....AspLysIleGluLysThrGlyLysIleLys 1042
||||| ..... |||||
2044 AGAAAGAGAGATTAAGCTGCAAAAGCTGAGCAAGAAAGTACTCTGAAA 2090

seq_name: /STD1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL08765
seq_documentation_block:
ID ABL08765 standard; cDNA: 5407 BP.
XX
AC ABL08765;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide seq ID NO 20777.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacological; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001: 2001WO-US09231.
XX
PR 23-MAR-2000: 2000US-191637P.
XX
PR 11-JUL-2000: 2000US-0614150.
XX
PA (PEKE ) PE CORP NT.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
DR P-PSDB; ABB64662.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
PS Claim 1; SEQ ID NO 20777; 21bp + Sequence listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent

```

capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL1840-ABL16175) and the encoded proteins (AB85737-AB872072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 5407 BP; 1382 A; 1436 C; 1482 G; 1107 T; 0 other;

alignment_scores: Quality: 1171.50 Length: 1656
 Ratio: 1.396 Gaps: 68
 Percent Similarity: 50.664 Percent Identity: 26.932

alignment_block: US-09-757-781-2 x ABL08765 ..

Align seg 1/1 to: ABL08765 from: 1 to: 5407

```

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||
134 ATGAAGATCCGCTGCTGCGACGCGATTCTGCTTCCCTGCG 183
17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
|||||
184 TTCGGCGCAAGCTCTGCTCGGATCTCTCAAGAGCGCCAGCGCGCT 233
34 yTrArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
234 ACATTCAGAGCGCGCGCAAG...CCCGATTCTCGGGTGCACCGCGCGCAT 280
51 LeuGlnHisGlyAspGlyGlyIleLeuAspLeuAspAspIleLeuCysAs 67
|||||
281 CTCACAG...ACACAGTCGGGCAATCCGATCCGATGATGTCGCGCA 327
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 83
|||||
328 CGTGGCGGAGATCGGGAGCAGATATTGGCGCATTTGATGACCCAGAGC 377
84 .....ProHisHisGlyValAspGlyThrSer...AlaSer 94
378 CCGATCCAGAGATTCGCGAAGAGCGCGAGATGAGCATCGGCGCATGCG 427
95 SerThrGlyThrGlnSerProGluIlePheGlySerGluLeuGlyThrAs 111
|||||
428 TCCGTTGGCAGCGGTTCCGCGCATCTTCGCGATCC...ACCA 471
111 pAsnValSerAlaPheGlnProTyrGln...AlaThrSerGluIleGluVal 127
|||||
472 CAGCGAGCGCCCACTGTCGCGGGGATCTCCACGCCACACATGAGG 521
127 AlThrProSer.....ValLeuArgAlaAsnMetProLeuHis 139
|||||
522 TCACCCACACACATCGGAGCGGATGGCTGAGCTGGAGTTCGATGATG 571
140 ValArgArgSerSerAspProAlaLeuIleGlyLeuSerThrSerValSe 156
|||||
572 GTGCGTCGACAGTGTGATCCCAATCTCTGCGC...TCGCTGAAGAGAGA 618
156 rAspSerAsnPheSerSerGluGluProSerArgLysAsnProThrArgT 173
|||||
619 GGGCAGTAC.....AAAGCT 635
173 rPserThrThrAlaGlyPhe.....Leu 180
|||||
636 GGTGGCGGCGGCTCCCACTACGCTGGCGGAGATTCGCCGAGAGCGCTG 685
181 LysGlnAsnThrAlaGlySerProLysThrCysAspArgLysLysAspG1 197

```

```

686 TTCTCTGACAAAGCCGGGTGGCCAGTTATCGCCACAGGGAGAGAGACA 735
197 uAsn.....TyrArgSerLeuProA 204
736 CGATCCCGCATCAGTAAAGAACAGCTGTCATCAACACACAGCCGCC 785
204 rGAspThrSerAsnTyrPheSerAsnGlnPheGlnArgAspAlaArgSer 220
786 ATGCTGCTAATGAGAGCTCTCTCCGCTGAAACCATCAGCGTTGGCCGA 835
221 SerLeuSerAlaSerHisProMetValGly.....LysTr 232
836 TCCGGTCGCCCTGTCATGATGCAATTCTTGCGGATGGCAATGCTACAGTG 885
232 pLeuGlnLysGlnLys.....GlnAspGluAspGlyThrGluGlu.... 245
886 GATGAGGCGACGCGAAMAATTCAAAATCAGCCCGCCAGCCAGCAGCAT 935
245 ..... 245
936 ATCAGCAGGTTTCATCATGCTGTGTCATAAAGCGTGCCTACTCC 985
246 .....AspAsnSerArgValGluProValGlyHisAla 256
986 AGCAAGTCTTGCACGAGGAGAGAGCAAGCAAGAGCCCTTGGGACAGGC 1035
256 aAspThrGlyLeuGlnHisIleProAsnPheSerLeuAspAspMetValL 273
1036 ATATGATTCATCAGGAGG.....AAGATGTGTGAAA 1067
273 yLeuValGluValProAsnAspGlyGlyProLeuGlyIleHisValVal 289
1068 TGCTCTGATCATCAACGATGACGATCGCTGCGGTGATCTGCTG 1117
290 ProPheSerAlaArgGlyArgThrLeuGlyLeuValLysArgLe 306
1118 CCGGACAAAGACCATGCGCGT.....GCATCTGTGGTCAACATGCT 1158
306 uGluLysGlyGlyLysAlaGlnHisGlnAsnLeuPheArgLysAsnAspC 323
1159 GGAAGCGGCGGACCGACCGAGAGAGAGAGACTG...CGTGGTGTGATC 1205
323 ySileValArgLysAsnAspGlyAspLeuArgAsnArgArgPheGluGln 339
1206 GCATTTGGAGATCATGATCAAGTAAATGAGCTACCGCAATCGCAG 1255
340 AlaGlnHisMetPheArgGlnAlaMetArgThrProIleIleTrpPheH 356
1256 GTTCAGAGACACTCGCGCGGCTTTGGAG..... 1285
356 sValValProAlaAlaAsnLysGluGlnTyrGluGlnLeuSerGlnSerG 373
1286 .....ACCTCAGAGTTGAGAGTGTGAGTGTGCGCGG 1319
373 LuLysAsnAsnTyrTyrSerSerArgPheSerProAspSerGlnTyrIle 389
1320 ATGCGAATCGC.....CGCAACAGCGT 1342
390 AspAsnArgSerValAsnSerAlaGlyLeuHisThrValGlnArgAlaPr 406
1343 GACTCTTAAGGTGGCCAGATGTGAGTGTGCTACGCTGCA..... 1384
406 oArgLeuAsnHisProProGluGlnIleAspSerHisSerArgLeuProH 423
1384 ..... 1384
423 tSerAlaHisProSerGlyLysProProSerAlaProAlaSerAlaPro 439
1385 .....CCACCCGCAACACATGCTGCTCGGAGGCGACCTCG 1423
440 GlnAsnValPheSerThrThrValSerSerGlyTyrAsnThrLysLysI1 456
|||||

```


860

```

3247 GTTTCGGATATATAACATGCCAAGATCGATCCAGTATCGAAGAAAG 860
860 luthrlysluasn...ThrValAspAspGlnLysAlaGlySerProSer 3296
3297 AGACGAGAGACAAATTGGCCCTGCGAGGCGGAGCGCAAGGATCAAGCTG 875
876 ArgAspValGlyProSerLeuGlyLeuLysSerSerSerLeuGly 3346
3347 GCGATCTGGGTCCATCGCTGGGAGCAAGAGTCTCTGCTGTGAGTCTC 892
892 lreuglnthrAlaValAlaGlnValThrLeuasnGlyAspIleProPhe 3396
3397 GCTCCAGACTATGCTGAGAGGCTGCAATGTCG...GATAGCGCGCTG 3443
909 lsaArgProArgProArgIleIleArgGlyArgGlyCysasnGlySerPhe 925
3444 GTCATCAGCGGTGGCGCAACCGCGTGGCGCTGGCAGGAGGACAGCTG 3493
926 ArgAlaAlaAlaAspLysSerTyrAsp...LysPro... 936
3494 CCGCGCGCGGTGGTGCAGCAACGATCGAGCAAGCCCGCAAGACCTG 3543
937 AlaValAspAspAspAspGlnGlyMetGlnThrLeuGlnGlnAspThrG 953
3544 GCTTTTGGAGATGCGCAT... 953
953 lugiLysSerArgSerGlyArgGlyLysValSerThrAlaSerAspGln 959
3576 TTGGCTCGCAGCGCAATGGA... 969
970 ProSerHisSerLeuGlnLysGlnMetAsnGlnLysGlnLysGlnLys 3595
3596 ... 986
986 plynThrAspArgLysLysAspLysThrGlyLysGlnLysLysAspA 3633
3634 CAAGTCGTCGGCGGCAAG... 3652
1003 rgaAspLysGlnLysAspLysMetLysAlaLysGlyMetLeuLysGly 3652
3653 ... 3673
1020 leuGlnLysPheThrPheArgPheGlyLysHisArgLysAspAsp 3673
3674 ATCGTACACTGTTCCGCTTGGCAAGATGCAAGATGCGTGGTGGC 3723
1034 ...LysIleGlnLysThrGlyLysIleLysIleGlnLys 3723
3724 AGTGATACATATGCGGTAAACATTTGCCCCCTACATCGGTGGTTCCA 3773
1046 erPheThrSer...GlnGlnGlnLysArgIleArgMet 1056
3774 CAGCCACATGCCCGCCAGCTGACAGCAGCAGCAGCAGCTGTCAGCAA 3823
1057 LysGlnGlnGlnLysArgIleGlnAlaLysThrArgLysPheArgLys 1073
3824 CATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1073
1073 g.GlnAlaArgLysArgAspTyrAlaGlnIleGlnAspPheHisArg 3873
3874 AAATGGCAAAACCGCGGTATGAC... 1089
1089 rPheGlyCysAspAspGlnLeuMetTyrGlyValAspSerTyrGlnG 3908
3909 CCGTGCCTGCTCC...AAATGGAGTGGTACGCAATGGTAT... 3945
1106 lYserMetAlaLeuAsnAlaArgProGlnSerProArgGlnGlyHisMet 1122
3946 ... 1122
1123 MetAspAlaLeuTyrAla...GlnValLysLysProArgLysSerTyr 1137

```

```

3964 CAACCATCGCTATCAGCATATATGCCAACTACAGAGACCTCCACAGAGC 4013
1137 sProSerProValAsp.SerAsnArgSerThrProSerAsnHisAspArg 1153
4014 ACCGAGCAACCGAGATTACTGCCCC... 1154
1154 lIleGlnArgLeuArgGlnGlnLysPheGlnAlaLysGlnAspGlnAsp 1170
4049 TACCATTCG... 1170
1170 lGlnAspArgArgArgThrTyrSerPheGlnGlnGlnProTyrProAsnAla 1187
4084 G...AGCAATGCATCGACAGAGCTCCGATCCAGCCAGCAGATCCCTGCTC 4130
1187 rgpProAlaThrGlnSer...GlyArgHisSerValSer... 4130
4131 AGCCGCAAGCGCAATCAAAAGCGCGTTCGCCATGAGCAGCTTACTACGAG 4180
1199 ...ValGlnValGlnMetGlnArgGlnGlnGlnGlnGlnGlnGlnGln 4180
4181 TAGGAGACGGTTCACAGCAAGCGGTGGCGGACATCAAGCATAGCAGCAG 4230
1214 rGlnGlnAlaGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysA 1231
4231 CAGCAGCGCC... 1231
1231 snAlaSerSerValSerGlnAspSerTyrGln... 1243
4251 CCTCGTCCCGCATCAATATGTCGCCATTTGGAGAGCGCGCAGCCATTAAGCC 4300
1244 TyrSerProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1260
4301 TACTGCGCAGCAGCCTGTAATAGCAGTCCCG... 4333
1260 rSerTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4333
4334 ... 4351
1277 laArgValMetLeuGlnThrGlnGlnLeuLeuArgGlnGln... 4351
4352 ... 1290
1291 ... 1291
4385 GCAATACCCCGCCAGCTGCTGACAGCAACATCAAGCAGCAAGCAGTGGC 4384
1300 sgInProProSer... 4434
4435 GCAGAGCCCGCCAGCTGACAGCTTCAAGAGATGTCAGCACTGCCAAT 4484
1310 yrasPserTyrLysLysValAlaAsp.ProSerTyrAlaProProLysGln 1326
4485 ATGGAGAT... 1326
1326 yProPheArg 1329
4526 TGACTATAGG 4535
seq_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.ABL08764
seq_documentation_block:
ID ABL08764 standard; cDNA: 36336 BP.
AC ABL08764;
XX
XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20774.
KW Drosophila; developmental biology; cell signaling; insecticide;
pharmaceutical; gene; ss.
XX

```



```
30132 .....CGCCAAACGCGTACTAGTGGCGGAGATGCT 30166
397 AGLYLEUHNISThValGlnArgAlaProArgLeuAsnHisProProGluG 30166
30167 GAGGGGGGTACGGCTGCA..... 30185
414 InitAspSerHisSerArgLeuProHisSerAlaHisProSerGlyLys 430
30186 ..... 30186
431 ProProSerAlaProAlaSerAlaProGlnAsnValPheSerThrTrpVa 30197
30198 CCAKATGCTGCTCCGGTGGGCACTCGCTGCAGTG..... 30233
447 IserSerGlyTrpAsnThrLysLysLysLysLysLysLysLysLysLysL 30233
30234 .....GCCAATACCGGTAACTGGGCAAGAAATCGAAATATATGC 30273
464 eulysLysGlyThnGluGlyLeuGlyPheSerIleThrSerArgAspVal 480
30274 TCAGAAAGGACCCCAAGCGCTGGGCTTTGGTCACAAACGGCGATTAAT 30323
481 ThrIleGlySerAlaProIleThrValLysAsnIleLeuProArgG1 497
30324 CCGCGCGGTGGCCACTGCTCCATCTACATCAAGATATCTGCCACGAGG 30373
497 yAlaAlaIleGlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluV 514
30374 TCGCGGCATCGAGATGAGCGCTGAAGCCCGGATCGTTCTCTCGAGG 30423
514 AlAsnGlyValAspLeuValGlyLysSerGlnGluValValSerLeu 530
30424 TGGATGGGACTCAATCACTGTGTAAACAAACGATGTGTGGCAATC 30473
531 LeuArgSerThrLysMetGluGlyThrValSerLeuValPheArgG1 547
30474 TTGAGGGGCATGCCAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 30523
547 n.....GluAspAlaPheHisProArgGluLeuAsnAla.... 558
30524 GAGGAGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 30573
559 .....GluProSer..... 561
30574 TGGCGGTGGCGGCTACGTGTGCTCACACAGCTGTCTCGACGTCAGCT 30623
561 ..... 561
30624 CCAGCGCTCCCATTCGGGTACAAAATCCAGACGACGACGATCTCTGT 30673
562 .....GlnMetGlnIleProLys..... 567
30674 CACTCATCAGCAGCAATCGCAGCTCAACGAAATCTCAGCACTTATGATG 30723
568 .....GluThrLysAlaGlnAspGluAspIleValLeuThrProAsp 581
30724 CGGCGACGAGTGGGGGCTCAAAATGATTAATCAATCAATCTTGA 30773
581 ..... 581
30774 GGAATTGGAACTATGCTATCTTTATGTCTTAAACACACAGACAGC 30823
582 .....GlyThr..ArgGluPheLeuThrPheG1 590
30824 CTGCGCGCCAGCAGCAGCAAGTTGGCACTCCCGGAGAGAGCTGACCTG 30873
590 uValProLeuAsnAspSerGlySerAlaGlyLeuGlyValSerValLysG 607
30874 CATTCCTGTACACACACGAAAGGCGGAGCTGGGTCTCAATGTGAAG 30923
607 LY..... 607
```

```
30924 GAAAGAGCTCTGAAATGTGAACGCTTCGGATGAGTGCCTCCAGCGGC 30973
608 ...AsnArgSerLysGluAsnHisAlaAspLeuGlyIlePheValLysSe 623
30974 AGCAATGAGACTGATGAGATGACATGACGCTACTGTTATTTCTGTGAAG 31023
623 rIleIleAsnGlyValAlaIleSerLysAspGlyArgLeuArgValAsn 640
31024 TCTAATTCATGGCGGTGCTGCTGATCCCGGATGCTGCTGGATGAATG 31073
640 sPGlnLeuIleAlaValAsnGlyLysLeuLeuGlyLysThrAsnGln 656
31074 ATCACTGCTGAGCGGTGAATGAGTATCGCTGGTGGGCAAAACAAATCC 31123
657 AspAlaMetGluThrLeuArgArgSerMet...SerThrGlyLysnly 672
31124 GAGCCATGGAGACACTACCTCGGGCAATGTGCMAACGCGCCGCAACA 31173
672 sArgLysMetIleGlnLeuIleValAlaArgArgIleSerLysCys.... 687
31174 TCGGGAGACCATTAACCTGTGTGTAGCGCCGCAAGATCTTGCAATCGCA 31223
687 ..... 687
31224 GTTCAGTGCATTTCTGGACCAAGTAACTACAGTATAGCATATAGC 31273
688 .....AsnGluLeu 690
31274 AATAGTAGCGGTGGCAGCAATTCMAATGTAGCGGCAATACAAATAG 31323
690 uLysSerProGlySerProProGlyProGluLeuProIleGlyThrAlaL 707
31324 CAGCTGAAATGCCAGCATTAATCTGGACCAACGGTCATATATTTGAATC 31373
707 euAspAspArgGluArgArg... 713
31374 CGGAGAGAGGAGGAGGAGCGCTCAATGGCGGAGGTGTGGCAGTGCT 31423
714 .....IleSerHisSerLeuTrp..... 719
31424 GGCAATGACATGAAATAGTGGGTAGTATTAATCTCATATATGTAGTACA 31473
719 ..... 719
31474 TAGTAACTAAGCTATTTGTCTTACAGATGAGCAATCCGTTTGAAT 31523
720 .....Ser 720
31524 CGCTAACCCTGGCATTTGCTCTCGAATTCAGCGACCATCTCTCA 31573
721 GlyIleGlu.....GlyLeuAspG1 727
31574 GCATCTCCACAGCAGCAGCGCATCTTCGACGACCAACAGCAGCAAC 31623
727 uSerProSerArgAsnAlaIleLeuSerArgIleMetGlyLysLysGly 743
31624 GTGCGCTGCTGAGCGCCGCTTTCAGCAGAGTGCAGCTGTAGAAAGAG 31673
744 LysTyrgLysLeuSerProThrValAsnMetProGln..... 755
31674 AGCTACTATATGACCACCAATGACAACTGTGCGCGGCGAGATGACATT 31723
756 .....AspAspThrValIleIleGluAspAspArgLeuP 767
31724 AATGACTGCTCATGCAACACAGCGCTGCTTATTCGAGGACGATGCCAGC 31773
767 roValLeuPro..... 770
31774 CGATGTCCGTGAGTGTGATAGTATATCTCAAAAGCAGCTTTTA 31823
771 .....ProHisLeuSerAspGlnSerSerSerSerSer.HisA 783
31824 ACTGACTTATTCCTATTTCTTTAGAAACACTACCGGACGCTCGCATG 31873
```

[illegible]

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saiko K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-18749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -

PS Claim 1; SEQ ID 3944; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

SO Sequence 556 BP; 154 A; 129 C; 144 G; 126 T; 3 other;

alignment_scores:

Quality: 936.00 Length: 184
Ratio: 5.200 Gaps: 0
Percent Similarity: 97.826 Percent Identity: 97.283

alignment_block:

US-09-757-781-2 x AAH07109

Align seg 1/1 to: AAH07109 from: 1 to: 556

126 gluValThrProSerValLeuArgAlaAsnMetProLeuHisValArgAr 142
|||
3 GAGGTACACCTTCAGTCTTCGAGCAAAATATGCTCTTCATGTTGACG 52
142 gSerSerProAlaLeuIleGlyLeuSerThrSerValSerAspSera 159
|||
53 CAGTAGTACCCACCTTAATGGCTTCACCTCTGACGTATAGTA 102
159 snpSerSerGluGluProSerArgValAsnProThrArgTrpSerThr 175
|||
103 ATTTTCTCTGAAGACCTTCAGAGGAAAAATCCACACGCTGGTCAACA 152
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAs 192
|||
153 ACAGCTGGCTCTCAGCAGAACACTCTGAGGTCCTAAAGCTGCCGA 202
192 PdgLysLysAspGluAsnTyrArgSerLeuProArgAspThrSerAsnT 209
|||
203 CAGGAGAAAGATGAAACCTACAGAACCTCCCGGAGATAGTAAGTAACT 252
209 rPserAsnGlnPheGlnArgAspAsnAlaArgSerSerSerSerAlaSer 225
|||
253 GGTCAACCAATTCAGAGACAAATGCTGCTCTGAGTCCAGT 302
226 HisProMetValGlyLysTrpLeuGluLysGlnGluGlnAspGluAspG1 242

|||||
303 CACCAATGTGGCAAGTGGCAGGAACAGACAGATCGAGATGC 352
242 yThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspThrG 259
353 GACGAGAGAGATTAACAGTCGTGTGAACCTGTGGACATGCTGCACACG 402
259 LysGluHisIleProAsnPhseSerLeuAspAspMetValLysLeuVal 275
403 GTTGGAGCATATACCAACTTTCTCTGATGATATGTAAGCTCTGA 452
276 GluValProAsnAspGlyGlyProLeuGlyIleHisValAlaProPhe 292
453 GAAGTCCCAACGATGAGAGGCTCTGGAATCCATGATGTCCTTCAG 502
292 rAlaArgGlyLysArgPheLeuValLysArgLeuGluLysG 309
503 TGTGCGANCGGCANNAACCTGCGTATATAGTAAACGATTGGGAAAG 552
309 Ly 309
|||
553 GT 554

seq_name: /SID51/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH71480
seq_documentation_block:
ID AAH71480 standard; cDNA: 597 BP.

AAH71480:

DT 19-SEP-2001 (first entry)

DE Human cervical cancer marker nucleic acid 2754.

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.

PN WO200142467-A2.

PD 14-JUN-2001.

PF 08-DEC-2000; 2000MO-US33312.

PR 08-DEC-1999; 99US-0169681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210600.

PR 21-JUL-2000; 2000US-0220114.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Deeds J, Berger A, Zhao X;

DR WPI; 2001-375006/39.

PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
PS Claim 1; Page 566; 1051pp; English.

CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition; to select
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy.

SO Sequence 597 BP; 180 A; 123 C; 156 G; 138 T; 0 other;

alignment_scores:

Quality: 902.50 Length: 209
Ratio: 4.725 Gaps: 1
Percent Similarity: 91.388 Percent Identity: 88.995

alignment_block:

US-09-757-781-2 x AAH71480 ..

Align seg 1/1 to: AAH71480 from: 1 to: 597

```
468 ThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyI 484
|||||
1 ACAGAAAGGTTGGGATTCACGATCACCCTCCAGAGATGTAACAATAGGTGG 50
|||||
484 YSerAlaProIleTyValLysAsnIleLeuProArgGlyAlaAlaIleG 501
|||||
51 CTACAGCTCCAAATCTATGTGTAACAAATCTCCCGGGGGGGCCCAATTC 100
|||||
501 LAspGlyArgLeuLysAlaGlyAspArgLeuIleGlyValAsnGlyVal 517
|||||
101 AGGATGGCCGACTTAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTA 150
|||||
518 AspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSerTh 534
|||||
151 GATTAGTGGGCAAAATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 200
|||||
534 rLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 551
|||||
201 CAAGATGGAAAGAGACTGTGAGCCTTCTGTCTTCCGCGAGAGAGAGCCCT 250
|||||
551 heHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLys 567
|||||
251 TCCACCCCAAGGGAACGTG..... 267
|||||
568 GluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrAr 584
|||||
268 .....AAACGAGAGAGATGAGATATTGTTCTTACACCTGATGGCACACAG 311
|||||
584 GGIupheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGlyL 601
|||||
312 GGAATTTCTGACATTTGAACTCCCACTTAATGATTCAGATCTGCAGGCC 361
|||||
601 euGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLeu 617
|||||
362 TTGGTGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 411
|||||
618 GlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
|||||
412 GGAATCTTTGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 461
|||||
634 YArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 651
|||||
462 AAGGCTTTGGTGAATGATCACTGATAGCAATTAATGAGAAATCCCTGT 511
|||||
651 euGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSer 667
|||||
512 TGGGCAAGACAAACCAAGATGCCCTGGAACCAACCTAAGAGGTATGTCT 561
|||||
668 ThrGluGlyAsnLysArgGlyMetIle 676
|||||
562 ACTTGAGGCCATTAACGAGAGATGATC 588
```



```

/ gene="PAR03"
/ note="Region: PDZ domain"
misc_feature      / gene="PAR03"
                  / note="Region: PDZ domain"
BASE COUNT      1220 a      923 c      1079 g      849 t
ORIGIN

alignment_scores:      quality: 7035.00      length: 1356
                       ratio: 5.188      gaps: 0
Percent similarity: 100.000      Percent identity: 100.000

alignment block:
US-09-757-781-2 x AFA67002 ..

Align seg 1/1 to: AFA67002 from: 1 to: 4071

1  MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
   |||||
1  ATGAAAGTACCGTGTGCTCGACGAGACCCGGGTGGTGGTCCGCGG 50
   |||||
17 yAspGlyHisMetLysValPheSerLeu1LegInGlnAlaValThrArgt 34
   |||||
51 GACCGCCACATGAAAGTTTCAGCTTCATCCACAGCGCGGTACCCGCT 100
   |||||
34 yTrArgLysAlaIleAlaLysAspProAsnYrTrP1LegInValHisArg 50
   |||||
101 ACCGGAAGGCCATCGCCAAAGATCCAAACCTAGTGAACAGTGCATCGC 150
   |||||
51 LeuGlnHisGlyAspGlyLysIleLeuAspLeuAspP1LeuLysCys 67
   |||||
151 TTGGAAACATGAGATGAGAAATACAGCTGTGATGACATTTCTTGTGA 200
   |||||
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
   |||||
201 TGTAGACAGCATTAAGACAGACTGTAGCAGTGTGATGACGAGATC 250
   |||||
84 roHisHisGlyLysPglyThrSerAlaSerSerThrGlyThrGlnSer 100
   |||||
251 CACATCAGGAGTGTGATGACACCGATCCAGTCCAGGGTACCCAGAC 300
   |||||
101 ProGluIlePheGlySerGluLeuGlyYrThrAsnAsnValSerAlaPheG1 117
   |||||
301 CCAGGATATTTGGTAGTGAAGCTTGACACCAACATGTCTACGCTTTCA 350
   |||||
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArgA 134
   |||||
351 GCCTTACCAAGCACACAGTGAATTTGAGTCCACACTTCAGTCCCTCGAG 400
   |||||
134 laAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
   |||||
401 CAAAATTCCTCTTCATGTTGCAGCAGTAGTACCCAGCTCTAATTGGC 450
   |||||
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
   |||||
451 CTCCTCCACTCTGTCAAGTATGTAATTTTCTCTGAAGAGCCCTTCAAG 500
   |||||
167 gLysAspProThrArgTrpSerThrAlaGlyPheLeuLysGlnAsnTr 184
   |||||
501 GAAAAATCCACACAGCTGTCAACACACAGCTGCTTCCCTCAAGCAGACA 550
   |||||
184 nAlaGlySerProLysThrCysAspArgLysLysAspGluAsnYrArg 200
   |||||
551 CTGCTGGAGATCTTAAACCTGACACAGAAAGAAAGTGAACCTACACA 600
   |||||
201 SerLeuProArgAspTrpThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
   |||||
601 AGCTTCCTCCGCGGATACAGTACCTGCTTCAACCAATTTCAAGAGACAA 650
   |||||
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
   |||||

```

```

651 TGTCTGCTCTCTCTGAGTGCACCTACCCCATGTGTGGGCAAGTGGCTGG 700
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
   |||||
701 AGAACAAGAACAGATGAGATGGACAGAAAGAGATTAACAGTGGTGT 750
   |||||
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
   |||||
751 GAACCTGTGGACATGCTGACAGCGGTTTGGAGCATATACCAACTTTC 800
   |||||
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
   |||||
801 TGTGATGATATGTGTAAAGCTCTAGAAAGTCCCAACAGATGAGAGGCC 850
   |||||
284 euGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300
   |||||
851 TGGGATTCATGTAGTGCTTTCAGTCTCGAGCGGCGCAACCCCTGGGG 900
   |||||
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLe 317
   |||||
901 TTTTATGTTAAACGATGTGAGAAAGTGTGTAACCTGACATGAAATTC 950
   |||||
317 uPheArgLysAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
   |||||
951 TTTTCTGAGAAATGATGTGATTCAGGATTAATGATGGCACCTTGAA 1000
   |||||
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
   |||||
1001 ATGAAAGATTTGAAACAGCAACACATATGTTTCGCAAGCCATCGTACA 1050
   |||||
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnYrG1 367
   |||||
1051 CCCATCATTTTGTTCCATGTGGTCTCGACCAAAATAAGACCATATAGA 1100
   |||||
367 uGlnLeuSerGlnSerGluLysAsnAsnYrTrpSerSerArgPheSerP 384
   |||||
1101 ACAATATCCCAAGTAGAAGAACAAATTCATCTCAAGCCGTTTATAGC 1150
   |||||
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
   |||||
1151 CTGACAGCCAGTATATTGACAAACAGAGTGTGAACATGAGGGCTTCA 1200
   |||||
401 yThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAsp 417
   |||||
1201 ACGGTCCAGAGAGCACCCGACTGACACACCCGCTGAGAGATAGATC 1250
   |||||
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
   |||||
1251 TCACCTAAGACTACCTCATAGCGCACACCCCTCGGGAACACCATCCG 1300
   |||||
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
   |||||
1301 CTCGAGCTCGGACCTCAGAAATGTTTATAGACACTGTAAAGCGTGT 1350
   |||||
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysG1 467
   |||||
1351 TATAACACCAAAAATAATAGCAGAGGCTTAATATCTACACTTAAGAA 1400
   |||||
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
   |||||
1401 TACAGAGGTTTGGGATTCACACATCACTTCAGAGATGAACATATGGTG 1450
   |||||
484 lYSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaIle 500
   |||||
1451 GCTCAGCTCCAAATCTATGTGAAAACATTTCTCCCGGGGGGGCGCAT 1500
   |||||
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGly 517
   |||||
1501 CAGGATGCGGCACTTAAGCAGAGACAGACTTATAGAGTTAATGAGAT 1550
   |||||
517 laAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSer 534
   |||||
1551 AGATTAGTGGCAAAATCCAAAGAGAAATGTTTTCGCTGTGAGAAACA 1600
   |||||

```

534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
 |||||
 1601 CCAGATGAGAGAACTGTGAGCCTTCTGTCTTTCGCCAGGAAGCGCC 1650
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnLeuProly 567
 |||||
 1651 TTCCACCCCAAGGAACTCAATCAGACGCCAGCCACATGCAGATTCCAAA 1700
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
 |||||
 1701 AAAAAAGAAAGCAGAGATGAGGATATGTCTTACCTGATGGCAGCA 1750
 584 rGluPheLeuThrPheGluValProLeuAsnAspSerLysSerAlaGly 600
 |||||
 1751 GGGAAATTTCGACATTGGAAGTCCCACTTAATGATTACAGATCTGCAAGC 1800
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
 |||||
 1801 CTTCGTGTAGTGTCAAGGTAAACCGGTCAAAAAGAACACCAGCAGATT 1850
 617 uGlyIlePheValLysSerIleIleAsnGlyAlaIleAspLysArgSg 634
 |||||
 1851 GGGAACTTTTGTCAAGTCCCATTTATATGAGAGAGCAGCATCTAAAGATG 1900
 634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
 |||||
 1901 GAAGGCTTGGGGTGAATGATCAACTGATGACGAAATGAGAAATCCCG 1950
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
 |||||
 1951 TTGGGGCAACAAACCAAGATGCATGGAACCCCTAAGAAAGCTATATGTC 2000
 667 rThrGluGluAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
 |||||
 2001 TACTGAAGCATTAACGAGATGATCCAGCTTATTTGTCAGAGAGAA 2050
 684 LeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700
 |||||
 2051 TTAAGCAAGTCATAGTAGTGAAGTCACTGGGAGCCCCCTGAGACTGAG 2100
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
 |||||
 2101 CTGCCCATTTGAAACACGCTTGATGATAGAGAAAGAAATTTCCCATTC 2150
 717 rLeuThrSerGlyIleGluGlyLeuAspGluSerProSerArgAspAla 734
 |||||
 2151 CCTTACAGATGGGATGAGGGGCTTATGATGATCCCCAGCAGAAATGCTG 2200
 734 LeuSerArgIleMetGlyGluSerGlyLysThrGlnLeuSerProThr 750
 |||||
 2201 CCTCAGTGAAGATTAATGGGTGAGTCAAGTAAATACAGAGTGTCCCTACA 2250
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
 |||||
 2251 GTGAATATGCCCCAGATGACACTGTCAATATGAAAGATGACAGTGGCC 2300
 767 oValLeuProHisLeuSerAspGlnSerSerSerSerSerHisAspA 784
 |||||
 2301 AGTGCTTCTCCACATCTCTGACCAAGTCTCTCCAGCTCCACGATGATG 2350
 784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
 |||||
 2351 ATGTGGGTTTGTGACGCGAGATGCTGTACTTGGGCCAAAGCTGCAATC 2400
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
 |||||
 2401 AGGTATTCAGCCGACATGCTTGTGAGTCCAGATGTGATCCAGTCTTCGC 2450
 817 apHeGlnArgGluGlyPheGlyLysArgLysMetSerGluLysArgThrL 834
 |||||
 2451 TTTTCAACGAAAGAAATTGGACGTCAAGATATGTCAGAAAAACGCAACA 2500

834 ySGLnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
 |||||
 2501 AGCAATTTTCAGATGCCAGTCATGATGATTCGTAAAAACAGCAAAATCA 2550
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
 |||||
 2551 AAAAGATGATTTAGTATAGCTGACGAGACTAAACTCAATACAGTGA 2600
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 |||||
 2601 TGACCAAGAAAGCAGGTTCTCCAGCAGAGATGTGGGTCTTCCCTGGGTC 2650
 884 eULysLysSerSerSerLeuGluSerLeuGlnThrAlaValAlaGluVal 900
 |||||
 2651 TGAAGAAGTCAAGCTCGTTGAGAGTCTGCAGACCGCAGTATGCCAGAGT 2700
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917
 |||||
 2701 ACTTTCAATGGGATATTCCTTCCATCCGTCCACGGCCCGGATAAATCAG 2750
 917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerThrA 934
 |||||
 2751 AGCCAGGGGATGCAATGAGAGCTTCAGAGCTGCCATGCAGAAATCTTATG 2800
 934 sPLeuProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950
 |||||
 2801 ATAAACCCCGGTAGATGATGATGATGAAGCATGAGACTTGGAAAGAA 2850
 951 AsPThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
 |||||
 2851 GACACAGAAAGAAATTCAACATCAGGAGAGAGTCTGTATCCACAGCCAG 2900
 967 rAspGlnProSerHisSerLeuGluArgGluMetLysGlnGlyAsnGlnL 984
 |||||
 2901 TGATCAAGCTTCCCACTCTCTGGAGAGACAAATGAATGAAACCAAGACA 2950
 984 ySGLysPlyThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
 |||||
 2951 AAGGTATGAAGTATGATAGAAAGAAAGATTAACCTGGAAGAAAGAAAG 3000
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
 |||||
 3001 AAAGATAGACATTAAGGAAGATTAATAAAGCCAAAGAAAGGAAGCT 3050
 1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034
 |||||
 3051 GAAGGCTTGGAGACATGTTCAGGTTGGCAACATCGAAAGATGACACA 3100
 1034 ySILeGluLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050
 |||||
 3101 AGATTGGAAGAAACGGGTAAATATAAATACAGGAATCTTTACATCGAA 3150
 1051 GluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysTh 1067
 |||||
 3151 GAGGAGAGATACGAAGAAAGACAGCAGAGAGAGATTCACAGCAAAAC 3200
 1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTrpArgIleGlu 1084
 |||||
 3201 TCGAGATTTTAAAGGAGACAGACTCGAGAGCGGACTATAGCTGAATATC 3250
 1084 InAspPheHisArgThrPheGlyCysAspAspGluLeuMetTrpGlyL 1100
 |||||
 3251 AAGATTTTCATCGGACATTTGGCTGTGATGATGATTAATGATGGGGA 3300
 1101 ValSerSerTrpGluGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
 |||||
 3301 GTTTCCTTATGAAAGGTTTCATGCGCTCTCAAGCCTGAGACCTCAGACCC 3350
 1117 oArgGluGlyHisMetMetAspAlaLeuTrpAlaGlnValLysLysProA 1134
 |||||
 3351 ACGAGAAAGGCATATGATGATGCTTTGTATGCCCAAGTCAAGAGAGCCG 3400
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150

```
|||||
3401 GGAATTCGCAACCTCAGCTGAGACAGTACAGTCACTCCTACCAAT 3450
1151 H1sAspArg11eGlnArgLeuArgGlnGlnGlnGlnGlnGlnGln 1167
3451 CATGATCGGATACAGGCTGTGAGGCAAGATTTCACCAAGCAAGCAAGA 3500
1167 pGluAspVal1GluAspArgArgArgThySerPheGlnGlnProTrp 1184
3501 TCAAGATGTAGAAGATCGTCGGGGACCTATAGTTTGAACAACCTGGC 3550
1184 roAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
3551 CGAAGCGACAGCGCGGAGCAGAGAGCGGCGACACGCTGTCGGGAG 3600
1201 ValGlnMetGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1217
3601 GTGCAGATCGACGCGGAGCGGAGAGAGAGCCGAGAGCTCCAGCAGGC 3650
1217 aGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSer 1234
3651 CCAGCGCCAGTACAGCTCTCGCTCGGCAAGAGGAAAAATGCCAGCT 3700
1234 eValSerLysAspSerTrpGlnGlnAspTyrSerProGlyGlyPhe 1250
3701 CGGTCCTCCAGGACTCTGGGAGCAGAACTACTCCCTGGGAGAGCTTC 3750
1251 GlnSerAlaLysGlnAsnProArgTyrSerSerTyrGlnGlySerArg 1267
3751 CAGAGTGCAGAAAGAGAACCCAGGTACTCAGCTACCAAGCTCCAGGAA 3800
1267 nGlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThr 1284
3801 CGGTCACCTGGGAGGAGCATGCTCAACGCCAGGGCAGTGTGAAACCT 3850
1284 InGluLeuLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1300
3851 AGGAGCTCCTTGCAGGAGAACAGGCGGAGAGACAGCAGAGTGAAGAAG 3900
1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysValGln 1317
3901 CAGCTCCTCGGAGGGGCCAGCAACTATGACTGATATAAAGTCCA 3950
1317 nAspProSerTyrAlaProProLysGlyProPheArgGlnAspValPro 1334
3951 GGACCCCAAGTTACGCCCTCCCAAGGCGCTTCCGGCAAGATGCCCC 4000
1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGlnLys 1350
4001 CCTCCCTTCTCAGGTTGCGAGGCTGAAACAGCTTCAAGCTCTGAGAAA 4050
1351 GlyArgProPheTyrSer 1356
4051 GGGAGGCCCTTCTATTCC 4068
seq_name: gb_pr:AF467003
seq_documentation_block:
LOCUS AF467003 4062 bp mRNA linear PRI 07-FEB-2002
DEFINITION Homo sapiens partitioning-defective 3 protein splice variant b
(PAR3) mRNA, complete cds, alternatively spliced.
ACCESSION AF467003
VERSION AF467003.1 GI:18568345
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4062)
AUTHORS Gao,L., Macara,I.G., and Joberty,G.
TITLE Multiple splice variants of PAR3 and of a novel related gene,
PAR3L, produce functionally different proteins
JOURNAL Unpublished
```

```
REFERENCE 2 (bases 1 to 4062)
AUTHORS Gao,L., Macara,I.G., and Joberty,G.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-2002) Center for Cell Signaling, University of
Virginia, PO Box 800577, Charlottesville, VA 22908, USA
FEATURES
source
1..4062
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..4062
/gene="PAR3"
/feature="PAR3"
1..4062
/feature="PAR3"
/feature="PAR3B: multi-PDZ domain containing protein;
alternatively spliced"
/codon_start=1
/product="partitioning-defective 3 protein splice variant
b"
/protein_id="FAL76043.1"
/db_xref="GI:18568346"
/translation="MKYTVCFGRTRVYVPCGGDHMKVFSILQDAVTRYKAIKDPNY
WIVVHRLHEDGGILDLIDLCDVADKRLVAFEDQDHHGDSLSASTGQDPE
IFGSELGDNVNSAFOPYQATSELEVTPSYLRAMMPLHYRSDPALIGSTVSDNSE
SSEPSKKNPTRMSTAGFLKONTAGSPKCDKDEYNSRLPRTSNNMNOFONRA
RSSLSASHPWVGKMLEKOBODEDGTEDNRSRVGHADTGLHINFSILDMDKTYE
VPNDGPGIGHVYVPSARGSRITGLYKYLEKGAKEHNTFRENCTYRINDGLRN
RRREQOAHFRQMRPTLTFHVVPANKDQYRQLSQSEKNNTYSRFSFSDSYIDNR
SVNSAGLHTVQAPRLNHPPEQIDSHSRPLPHSNAPSGKPPSAPASAPQVNFSTVBSG
YNRKKGIRLNIQLKKTGELGFSITSRDVTIGSNAPYVYKNTLIPRGAALIDSRRLAG
DRLIEVNGVDLVKSOEYVSLRSMETVSLVFEQDPAHRELNAPESQOMIP
KERTKADEDVILPDGTREFLFEVPLNMGSGSLGVYGNRSKKNHMDLGFVFSI
INGGASKIDGRLLRVNDOLIAVNGESILGKTNDQAMTILRSMSTBEKNGMQLIYAR
RISKCMETKSPGSPGPELPIETALDDBRRIRSHSYSGIEGIDESPKNALSRITG
KYDSPTVYMPQDDTYITEDRLPLVLPHLSDSSSSHDDGFTVADGVTAKAAS
DSADCSLSDVDVYLAIFQREGFGQSMSEKRTKQFSDASQLDFVTKRKSSMDLGIAD
ETKNTLVDDQKAGSPRDVGPISGLKSSLSLSIQTAVAEVTLNGDIPIHRRPRLIR
GRGNSFSRAIDKSYDKPAVDDDDGEMTELEDTESSRSRGSYSTASDSSHLE
ROMNGOENQKDDTKDRKKDKTKGKKDKRDKKDKKAKKGMKLGIDMFRGHRRD
KIEKTKIKIOESFTSEERRIRMKROBERTOAKTRFRERROARERYAETIODHRTFG
CDEELMAYGVSSFEQSMALNARPOSRECHMDALTAQYKPKPNSKPSFVDSNRSTPS
NHDRIORLROEFOQAOODEVEDERRTTYSEDDWPAPARATYSGRHSVSEVQMOQR
QEERESSQOARQYSSLPKRSKSNASVSODSEWQYSGDEGOSAKENPNSYSYGS
RMNGYLGHGHNARVMLETOELRQEOERKQKOPPEGSPSNYSYKKVODPSYAP
PKGPFRODVPSPSGVAVARLRLQTPKGRPFYS"
814..1074
/gene="PAR3"
/feature="PAR3"
/feature="PAR3"
1378..1638
/gene="PAR3"
/feature="PAR3"
1765..2028
/gene="PAR3"
/feature="PAR3"
/feature="Region: PDZ domain"
/feature="Region: PDZ domain"
/feature="Region: PDZ domain"
BASE COUNT 1218 a 922 c 1075 g 847 t
ORIGIN
```

alignment_scores:

Quality:	7008.50	Length:	1356
Ratio:	5.180	Gaps:	1
Percent Similarity:	99.779	Percent Identity:	99.779

alignment_block:

US-09-757-781-2 x AF467003 ..

Align seg 1/1 to: AF467003 from: 1 to: 4062

```
1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||
1 ATGAAAGGACCGCTGTGTTGGAGACGACCGGCTGTGTCGCGCGG 50
17 yAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
```

```

|||||
51 GGAGGGCCCATGAAAGTTTTCAGCCCTCATCCAGCAGCGGTACCCGCT 100
34 yfArqYsAlaIleAlaYsAPProAsnTyTrpIleGlnValHisArg 50
|||||
101 ACCGGAAGCCATCGCCCAAGATCCAAACTACTGATACAGTGCCATCGC 150
51 LeuGlnHisGlyAspGlyLysIleLeuAspLeuAspPrlLeuGlyAs 67
|||||
151 TTGGACATGAGATGAGGAATAGACTGATGACCTGATGACATCTTGTTGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValAlaPheAspGluGlnAsp 84
|||||
201 TTGAGAGAGAGATAAAGACAGACTGGTAGCTGTTGATGAGACAGATC 250
84 roHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
251 CACATCACGAGAGGTGATGGCACACAGTGCCAGTTCCACGGGTACCCAGAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG 117
|||||
301 CCAGAGATATTTGTGTAGTGTGGCACCAACAATGTCACAGCTTTTCA 350
117 hProTyTrpGlnAlaThrSerGluIleGlnValAlaThrProSerValLeuArg 134
|||||
351 GCCTTACCAAGCAACAAGTGAATGAGGTCAACACCTTCAGTCCCTTCGAG 400
134 lAsnMetProLeuHisValArgArgSerSerProAlaLeuIleGly 150
|||||
401 CAAATATGCCCTTCATGTTGCGAGCATGATGACCAAGCTCTTAATTTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
|||||
451 CTCTCCACTTCTGTCAGTGATGATTAATTTTCTCTGAAGAGCCTTCAAG 500
167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsn 184
501 GAAAAATCCCAACAGCTGTCAACAACAGCTGCTTCCTCAACACAGACA 550
184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyArg 200
|||||
551 CTGCTGGAGACTCTTAAACCTCGCAGACAGAAAGATGAAACTACAGACA 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
601 AGCCTCCCGGGGATAGTACTAGTAACTGCTTAACCAATTTCCAGAGACAA 650
217 naIAsnArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeu 234
|||||
651 TGCCTCGCTGCTCTGAGTCCAGTCAACCCAATGGTGGCAAGTGGCTGC 700
234 lLysGlnGlnGlnAspGluAspGlyThrGlnGlnLysAspAsnSerArgVal 250
|||||
701 AGAAACACAGACAGATGAGGATGGGACAGACAGATTAACACTCGTGTT 750
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
751 GAACGTTGTGACATGCTGACAGGGTTTGGACATATACCACAACCTTTC 800
267 rLeuAspAspMetValLysLeuValGlnValAlaProAsnAspGlyLysPro 284
|||||
801 TCTGGATGATGATGTAAAGCTCTAGAAAGTCCCAACAGATGAGAGGCTTC 850
284 euGlyIleHisValAlaProPheSerAlaArgGlyGlyArgThrLeuGly 300
|||||
851 TGGGATTCCTGATGATGCTTTCAGTCTCTGAGCGGCGCAGAACCTTGGG 900
301 LeuLeuValLysArgLeuGlyLysGlyGlyLysAlaGlnHisGlnAsnLe 317
901 TTATTGTGTAACGATTTGAGAAAGGTGTAAAGCTGAACATGAAAAATCT 950
317 upHeArgGluAsnAspCysTlleValArgIleAsnAspGlyAspLeuArg 334
|||||
|||||
951 TTTTCTGAGAAATGATTGTCAGGATTAATGATGGCGACCTTCGAA 1000
334 snArgArgPheGlnGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
|||||
1001 ATAGAAATTTGAAACAAGCAACATATGTTTGGCCAAAGCCATCGGACA 1050
351 ProIleIleTrpPheHisValAlaProAlaAlaAsnLysGluGlnTyrg 367
|||||
1051 CCCATCTATTGGTTCATGTGTGTTCTCGACAGCAAAATTAAGACAGATGTA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyTrpTrpSerSerArgPheSer 384
1101 ACAACTATCCCAAGGTGAGAAACAATTAATTAATCAAGCCGTTTAAGCC 1150
384 roAspSerGlnTyTrlLeaAspAsnArgSerValAsnSerAlaGlyLeuHis 400
|||||
1151 CTGACAGCCAGTATATTGACACAGAGTGTGAACAGTGCAGGGCTTCAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1201 ACGGTGACAGAGACACCCGACATGAAACCCGCGCTGACAGATAGACTC 1250
417 rHisSerArgLeuProHisSerSerAlaHisProSerGlyLysProProSer 434
1251 TCACTCAAGACTACCTCATAGCGCACACCCCTCGGGAACCAACCATCCG 1300
434 lArProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1301 CTCACGCTCGGCACCTCAAAATGATATTGATAGACTGTAAAGCATGGGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1351 TATTAACACCAAAAAAATAGCAGAGAGCTTAATATCCAGCTTAAGAAGG 1400
467 yThrGlnGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
1401 TACAGAAAGTTTGGGATTCAGCATCACTTCCAGAGATTAACAATTAAGTG 1450
484 lYSerAlaProIleTyValLysAsnIleLeuProArgGlyAlaAlaIle 500
1451 GCTCAGCTCCAAATCTATGTGAAAAACATTTCTCCCGGGGGGCGGCATT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyVa 517
1501 CAGATGCGCGACTTAAGCAGAGACAGACTTATAGAGTAATGAGAGT 1550
517 lAspLeuValGlyLysSerGlnGlnGlnValAlaSerLeuLeuArgSer 534
1551 AGATTTAGTGGCAAAATCCCAAGAGGAAGTTGTTCCGCTTGTGGAAGCA 1600
534 hLysMetGlnGlyThrValSerLeuLeuValPheArgGlnGlnAspAla 550
1601 CCAAGATGGAAGAACTGTGAGCTTCTGTGTCTTCGCCAGGAAGAGGCC 1650
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProly 567
1651 TTCACCCCAAGGAATGAAATGACAGACCAAGCAGATGCGAGATCCAAA 1700
567 sGlnThrLysAlaGlnAspGluAspPrlLeuValLeuThrProAspGlyThra 584
1701 AGAAACGAAAGCAGAAAGATGAGATATTGTTTACACCTGATGGCACCA 1750
584 rGluPheLeuThrPheGlnValAlaProLeuAsnAspSerGlySerAlaGly 600
1751 GGGATTTTCGACATTTGAAGTCCACTTAATGATTCAGATTCGACAGCC 1800
601 LeuGlyValSerValLysGlyAsnArgSerLysGlnAsnHisAlaAspLe 617
1801 CTGGTGTCAGTGTCAAAAGGTAAACGGGTCAAAAGAGAACCAACGAGATT 1850
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
1851 GGGATCTTTGTCAAGTCAATTATTAATGAGAGAGACAGCATCTAAAGATG 1900

```

634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyIleSerLeu 650
 1901 GAAGGCTTCGGGTGAAATGATCACTGATAGCAGTAATGAGAAATCCCTG 1950
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
 1951 TTGGGCAACAAACCAACAGATGCCATGGAAACCTTAAGAAAGCTTATGTC 2000
 667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
 2001 TACTGAGGCATTAACGAGGATGATCCAGCTATTGTCTCAGAGGAA 2050
 684 IeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700
 2051 TAAGCAAGTCCAAATGAGCTGAACTCACCTGGGAGCCCTCGACCTGAG 2100
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
 2101 CTGGCCCATGAAACGCGTTGATGATAGAAACGAAATTTTCCATTTC 2150
 717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
 2151 CCTTACGCTGGGATGAGGGGCTTGATGAATCCCAAGCAGAAATGCTG 2200
 734 IeLeuSerArgIleMetGlyIleSerGlyLysTyrGlnLeuSerProThr 750
 2201 CCTCAGTAGGATATG.....GGTAAATACCAAGCTGTCCCTAC 2241
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
 2242 GTGAATATCCCAAGATGACACCTGTCATTTAGAAAGATGACAGGTGCC 2291
 767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAsp 784
 2292 AGTGCTCTCCACATCTCTGACCAGTCCCTTCCAGCTCCCATGATG 2341
 784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800
 2342 ATGGGGGTTGTGACGCGCAGATGCTGACTGGGGCCCAAGGCTCCAAATC 2391
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
 2392 AGTATTCAAGCGGACCTGCTCTTGAAGTCAGATGTGATCCAGTTCTTC 2441
 817 aPheGluArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
 2442 TTTTCACGAAAGAGATTGGAGCTCAGAGTATGTCAGAAAAACGACAA 2491
 834 ySgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
 2492 AGCAATTTCAGATGCCAGTCAATTGGATTTCCTTAAACACCAAAATCA 2541
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
 2542 AAAAGATGATTAGTATAGTACGACGACTAAACTCAATTAACAGTGA 2591
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 2592 TGACCCAGAAACAGGTTCTCCAGCAGAGATGTGGTCTTCCCTGGGTC 2641
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
 2642 TGAAAGAGTCAAGCTCGTTGAGAGCTTCAGAACCGCAGTTGCCAGGTG 2691
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleLeu 917
 2692 ACTTTGATGGGATATCTTCCATCGTCCACGCGCGCGATATATCAG 2741
 917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrA 934
 2742 AGCGAGGGGATGCAATGAGAGCTTCAGAGCTGCATCGAACAAATCTTATG 2791

934 sPLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGlu 950
 2792 ATAAACCCCGCGGTAGATGATGATGAGGATGAGACCTTGGAAAGA 2841
 951 AspThrGluGlnSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
 2842 GACACAAAGAAAGTTCCAGATCAGGAGAGAGCTGTATCCACAGCCAG 2891
 967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlnAsnGlnGlu 984
 2892 TGATCAGGCTTCCACTCTCTGAGAGACAAATGAAATGAAACCAAGAGA 2941
 984 ySgIleAspLysThrAspArgLysLysAspLysThrGlyLysGlyLys 1000
 2942 AAGGTGATTAACCTGATAGAAAAAAGATTAACCTGAAAAAGAAAGAG 2991
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
 2992 AAAGATAGAGATTAAGAGAAAGATTAATTAAGCCAAAGAAAGGATGCT 3041
 1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034
 3042 GAAGGCTTGGGAGACATGTTGAGGTTGGCAACATCGAAAGATGACA 3091
 1034 ySIIleGlyLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050
 3092 AGATTGAGAAACGGGTAAATTAATTAACAGATCCTTTACATCAGAA 3141
 1051 GluGluArgIleArgMetLysGlnGlnGlnGluArgIleGlnAlaLysThr 1067
 3142 GAGAGAGAGATTAAGAAAGCGAGGAGAGAGATTCAGCAAAAC 3191
 1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIleG 1084
 3192 TCGAATTTTAGGGAACACACACTCGAAGGCGCTGACTGCTGAAATTC 3241
 1084 IAspPheHisArgThrPheGlyCysAspAspGluMetLysGlyGly 1100
 3242 AAGATTTTCATCGGACATTTGGCTGTGATGATGATGATTAATGTTGGGGA 3291
 1101 ValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
 3292 GTTCTCTTATGAAGGTTCCATGGCTCTCAAGCCTAGACCTCAGAGCC 3341
 1117 oArgGluGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysPro 1134
 3342 ACGAGAAAGGCATATGATGATGATGCTTTGTATGCCAAGTCAAGAAAGCCG 3391
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
 3392 GGAATTCCAAACCTCTACCTGTAGACGTAAACAGATCAACTCTAGCAAT 3441
 1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
 3442 CATGATCGATACAGCGCTGAGGCAAGAAATTCAGCAAGCAAAACAGGA 3491
 1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProCtPP 1184
 3492 TGAAGATGTAGAAAGATGCTGGCGGACCTATAGTTTGAACAACCTGGC 3541
 1184 rAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
 3542 CGAAGCGACGCGCGCGACCGAGAGCGGGCGACACTCGGTGCCGTGGAG 3591
 1201 ValGlnMetGlnArgGlnArgGlnGluGluGluGluLysSerGlnGlnI 1217
 3592 GTGCAATGACAGCGGCAAGCGCAGAGAGCGCAGAGCTCCAGAGAGG 3641
 1217 aGlnArgGlnTyrSerSerLeuProArgGlnSerArgLysAsnAlaSer 1234
 3642 CCAGCGCCAGTACAGCTCTGCTGCGCAAGACAGGAAAAATGCCAGCT 3691
 1234 eValSerGlnAspSerTrpGluGlnAsnTyrSerProGlyGluGlyPhe 1250

|||||
3692 CGGTCTCCGAGGACTCTGGACAGAACTACCTCCCTGGGGAAGGCTTC 3741
1251 GlnSerAlaIysGlnAspProArgTyrSerSerTyrGlnIysSerArgas 1267
3742 CAAAGTCCCAAGAAAGAACCCAGTACTCCAGTACCAAGGCTCCAGAA 3791
1267 ngIyTyrLeuGlyGlyHISgIyPheAsnAlaArgValMetIeuGlnThrG 1284
3792 CGGCTACCTGGAGAGACATGCTTCAACGCCAGGATCATCTCGAAATC 3841
1284 InLleuLeuAlaArgGlnGlnIleuArgArgIysGlnGlnIleuIysLys 1300
3842 AGGAGCTCTCCGACAGAACAGAGCGGAAAGACGACGATGAAAGAG 3891
1301 GlnProProSerGlnIyProSerAsnTyrAspSerTyrIysLysValG 1317
3892 CACCTCTCTCCGAGGAGGCCACACTATGACTCGATTAAGAAATGCCA 3941
1317 naapProSerTyrAlaProProIySgIyProPheArgGlnAspValProp 1334
3942 GGACCCAGATTACGCCCTCCCAAGGGCCCTCCGCGCAGAGATGTGCC 3991
1334 roserProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGlnLys 1350
3992 CCTCCCTCTCAGGTTCGAGGCTGAACAGCTTCAGACTCTCGAGAAA 4041
1351 GlyArgProPheTyrSer 1356
4042 GGGAGGCCCTTCATTC 4059
seq_name: gb_pr:AF196185
seq_documentation_block: 5958 bp mRNA linear pri 16-NOV-2001
LOCUS AF196185 5958 bp mRNA linear pri 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long
variant mRNA, complete cds.
ACCESSION AF196185
VERSION AF196185.1 GI:13491609
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 5958)
Fang,C.M. and Xu,Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
asp isoforms in human hepatocellular carcinomas
JOURNAL Cell Res. 11 (3), 223-229 (2001) *SEPT*
MEDLINE 21520080
PUBMED 11642408
REFERENCE 2 (bases 1 to 5958)
AUTHORS Fang,C. and Xu,Y.
TITLE Exon/Intron Structure and Splicing Variants of a Novel Human
Polarity Gene, hASIP
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5958)
AUTHORS Fang,C. and Xu,Y.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Laboratory of Molecular and Cellular
Oncology, Shanghai Institutes for Biological Sciences, Chinese
Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P.R. China
FEATURES
source
1..5958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
288..4349
/note="hASIPa"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein long variant"

/protein_id="AAK27891.1"
/db_xref="GI:13491610"
/translation="MKVYCFGRTRVVPCCGDGHMKVFSLIQAAVTRYRKALADPNY
WIOVRLEHGGGIIIDLDILCDVADDDKRLVAVDEODPHHGGGTSSNSGQSP
IFGSRIGTNVNSARPOYCATSRILEYTPSVLRANMLHYRRSSDPALIGISTVSNSP
SEEPSRKPTRWSTTAGTGLKNTKASPTCDKDKDENTRSLPRTSMNSQFQDNA
RSSLSASHPWGWLEKQEDDGTEDNSRVEPVHADTGLEHLPNESLIDMKVLYE
VNDGPPILIHVPPSARGRTGLVLEKLEGEAEHNLRENDCIYRINDMDLNR
RRFEQAQHMFRQAMRTPLIWFHVVAANKEOYEOSREKNYYSRSFSDQYIDNR
SVNSAGLHVQAPRLNHPPEQIDSHSRLPHSAHPGKPSAPASAPQVFTVSSG
YMKKIKGRNLNOLKCTGEGLFSTSDPTGCSAPLYKNILPREGAIDGRKAKG
DRLIEYNGVDLYGKSQEEVYSLRSTKMGVSLVFRQEDLPHRELAEPSQMDIP
KETKADEDIVLTPOGTREFLFEVPLNDGSAAGLYSVKGRSKENHADLGIYKSI
INGGASAKSGRLRVNDOLIVNGESILGKTNDAAETLRSMSTEGNKRMQOLIYAR
RISKNEKSPGSPGPELPETALDDREERLISHLSGILEGDSPSKNALSTRMG
KYOLSPYVMPODDVLIIEDRLPVLPPLSHOSSSSSSHDGFPVTAADGAKAAIS
DSADSLSTFVDVPVLAFFQREGGRQSMSEKRTKQVSDASQDLFPVTRKSKMDLIAD
ETKLTIVDQKAGSPSRDVPGLGKSSLSLELQTAASVLELNDIIPHRPRLIR
GRGNESEFRALADKSYDKPAVDDEGMETLEDETESSRSGESVSTASDPSHLE
ROMNNOERGDYDKDKTKGKKKDKRDKKAKKAKGMLKGLDMRFGRKHKRD
KLEKTGKIKIOEFTSEERIRMKOBERIOAKTRFERROARDYARIDOFHRTFG
CDEELMYGVSSYEGSMALNARPOSPREGHMDALYAQVKPRNSKPSVDNSRSTPS
NHDRIORLROEFOAKODEVDVDRRTYSFRQPPNPARATQSGHSVSVEWOMOR
QEERSSQQAQOYISLPSROSKNMSVSQDSMEDONTSEBGFQSKENRPTSTQGS
KNGYLAGEFNARVMLETQELLROEORRQDQMKQPSBEGSVNSDYKRYDDPSYAP
PKGPRQDVPSPSOVARLNLRLQTEPKGRPFS"
BASE COUNT 1623 a 1395 c 1581 g 1358 t 1 others
ORIGIN
alignment_scores:
Quality: 7008.50 Length: 1356
Ratio: 5.180 Gaps: 1
Percent Similarity: 99.779 Percent Identity: 99.779
alignment_block:
US-09-757-781-2 x AF196185 ..
Align seg 1/1 to: AF196185 from: 1 to: 5958
1 MetLysValThrValCysPheGlyArgThrArgValValProGlyG 17
|||||
288 ATGAAAGTGAACCGTGTCTTGAGGAGACCCGGGTGGTGGCGTGGG 337
17 yaspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
|||||
338 GGAGGCGCACATGAAGTTTTCACCTCATCCACAGCGGTGACCCGCT 387
34 yThrGlyAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
388 ACCGGAAGCCCATGCCAAGATCTCAACTGATACAGGTGCATCGC 437
51 leuGlnHisGlyAspGlyIleLeuAspLeuAspAspIleLeuGlyAs 67
|||||
438 TTGAAACATGGAGATGAGGAATACTAGACTTGATGACATCTTTGGA 487
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnAspP 84
|||||
488 TGTAGCAGACGATTAAGACAGACAGTGTGATGATGAGCAGATC 537
84 roHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
538 CACATTCACGGAGGTGATGGCACACAGTCCACGCGGATACCAGAGC 587
101 ProGlnIlePheGlySerGlnLeuGlyThrAsnAsnValSerAlaPheG 117
|||||
588 CCAAGATATTGTGATGAGCTTGACCAACATGCTCTAGGCTTTCA 637
117 nProTyrGlnAlaThrSerGlnIleGlnValThrProSerValIleuArg 134
|||||
638 GCGTTACCAAGCAACAGTGAATTTGAGGTACACCTTCATCTTGAG 687
134 laaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150

|||||
688 CAAATATGCTTCTTCAATGTTCCAGCAGCATGATGACCCAGCTTAATGTC 737
151 LeuSerThrSerValSerAspSerAsnPhseSerGluLupProSerAr 167
|||||
738 CTCTCCACTTGTGTGAGTATGATATTTTCTCTGGAAGAGCTTCAAG 787
167 glysasnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
788 GAAAAATCCACACGCTGGTCAACAACACCTGGCTTCTCAAGCAGAACAA 837
184 hralaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrlarG 200
838 CTGCTGGAGATCTCTAAACCTGGACAGACAGAAAGATGAAAACCTACAA 887
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
888 AGCTTCCCGGGGATCTAGTAACTGTCTAACCAATTTTCAGAGAACAA 937
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
938 TGTCTCGTCTGTCTGAGTGCAGTCCACCAATGGTGGCAAGTGGCTGG 987
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
988 AGAAACAAAGACAGATGAGATGGGACAGAAAGAGATAAAGTGTGTT 1037
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPhse 267
1038 GAACTGTTGGACATGCTGACAGGGTTGGAGCATATACCCAACTTTTC 1087
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysIleProL 284
1088 TCTGGATGATATGCTAAAGCTCTCTACAAGTCCCAACGATGGAGGCTTC 1137
284 euGlyIleHisValAlaProPhseSerAlaArgGlyLysArgThrLeuGly 300
1138 TGGGAATTCATGTAATGCTTTCAAGTCTCGAGGCGGACAGAACCTGGGG 1187
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnIle 317
1188 TTTTGTAGTAAACGATTTGGAAGAGGTGTGAAGCTGAACATATAAATCT 1237
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
1238 TTTTCTGGAATGATTCATTTCTCAGGATTAATGATGGGACCTTCGAA 1287
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1288 ATAGAGATTTGAAACAAGCACAACATATGTTTCCCAAGCCATGCGTACA 1337
351 ProIleIleTrpPheHisValAlaProAlaAlaAsnLysGluGlnTyrl 367
1338 CCCATATTTGGTCCATGTGTCTCCGACGCAATTAAGCAAGCATATGA 1387
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrlTyrlSerSerArgPhseSer 384
1388 ACAACATATCCCAAGTGAGAAACAATTAATCAAGCCGTTTTATGCC 1437
384 roAspSerGlnTyrlIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1438 CTGACAGCCAGTATATTGACMACAGAGTGTGAACAGTGCAGGGCTTCAC 1487
401 ThrValGlnArgAlaProArgLeuAsnHisProGluGlnIleAspSe 417
1488 ACGGTGAGAGACACCCGAGCTGAACACCCGCTGAAGCATAGACTC 1537
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1538 TCACTCAAGACTACCTCATAGCGACACCCCTCGGGAACCAACCATCCG 1587
434 lAProAlaSerAlaProGluAsnValPheSerThrThrValSerSerGly 450
|||||

1588 CTCCAGCTCGGACCTCAGATGATTTAGTACGACTGTAAAGCATGCT 1637
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnIleLysLysG 467
1638 TATAACACCAAAAAATAGCAGAGGCTTAATATTCAGCTTAAGAAAGG 1687
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1688 TACAGAAAGTTTGGGATTCAGCATCATTCCAGAGATGTAAACATAGGG 1737
484 ySerAlaProIleTyrlValLysAsnIleLeuProArgGlyAlaAlaIle 500
1738 GCTCAGCTTCCAACTATGTGAATAAATTTCTCCCGGGGGGCGGCAAT 1787
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVA 517
1788 CAGGATGGCCGACTTAAGGAGAGACAGACTTATAGAGTAAATGAGAT 1837
517 lAspLeuValGlyLysSerGlnGluValAlaSerLeuLeuArgSerT 534
1838 AGATTTAGTGGCAATCCCAAGAGAAAGTGTTCGCTGTGAGAGACGA 1887
534 hrlYsMetGluGlyThrValSerLeuLeuValPheArgGlnLysPAla 550
1888 CCAGATGGAGAGAACTGTGAGCTTGTGCTTCCGACGAGAAACGCGC 1937
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
1938 TTCCACCCAAAGGAACTGAATCAGACGCCAACCCAGATGCCAATTCCAA 1987
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
1988 AGAAACGAAGACGAAGATGAGGATATTTGTTCACTACCTGAGCACCA 2037
584 rGluPheLeuThrPheGluValProLeuAsnAspSerLysSerIleGly 600
2038 GGAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGGATCTCCAGGC 2087
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspIle 617
2088 CTTCGTGTGATGTCACAAAGTAAACCGTCAAAAAGAAACACGACGATTT 2137
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
2138 GGAATCTTTGTCAAGTCCATTAATGAGAGAGACGACATCTAAAGANG 2187
634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 650
2188 GAGGGCTTCGGGTGAATGATCAACTGATAGCAGTAATGGAATTCCTG 2237
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
2238 TTGGGCAAGCAAAACCAAGATGGCATGGAACCCCTAAAGGCTATGTC 2287
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
2288 TACTGAAGCAATAAACGAGAAATGATTCACGCTTAATTTGGCAAGGAGA 2337
684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
2338 TAAAGAAAGTGAATGAGCTGAAGTCACTGGAGAGCCCTTGACCTGAG 2387
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
2388 CTGCCCCATTGAACACAGCTTGATGATAGAGAAAGAAATTTCCCATTC 2437
717 rLeuTyrlSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
2438 CCTCTACAGTGGGATTTAGGGGCTTGATGATGATGCCACAGAAATGCTG 2487
734 lAsuSerArgIleMetGlyLysSerGlyLysTyrlGlnLeuSerProThr 750
2488 CCTCTCAGTAGGATTAATG.....GTTAAATACAGAGTGTCCCTTACA 2528

751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
|||||
2529 GTGAATATGCCCCAAGATGACACTGTCATTATGAGATGACAGGTTGCC 2578
767 oValLeuProProHIsLeuSerAspGlnSerSerSerSerSerHIsAspA 784
|||||
2579 AGTGCTCTCCACATCTCTGTACACAGTCCCTTCCAGCTCCACATGATG 2628
784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
|||||
2629 ATGTGGGCTTTGTGACGCGACAGATGCTGTACTTGGGCCAAGCGTCGAATC 2678
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
|||||
2679 AGTGATTACAGCCGACTGCTCTTGTAGTCCACGATGTTGATCCAGTCTCTGC 2728
817 alpHeGlnArgGluGlyPheGlyValArgGlnSerMetSerGluLysArgThrL 834
|||||
2729 TTTTACACAGAGAGATTGGACGTGAGATGATCAGAAAAAGCAGACAA 2778
834 ysgGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
|||||
2779 AGCAATTTTCAGATGCCAGTCAATTGGATTTCGTTAAAAACAGCAAAATCA 2828
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
|||||
2829 AAAACGATGATTTAGTATAGCTGACGAGACTAAACTCAATACAGTGA 2878
867 paspGluLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
|||||
2879 TGACCGAGAAAGCGTTCTCCAGCAGAGATGGGGTCTCTCCCTGGGTC 2928
884 eulLysSerSerSerSerLeuGluSerLeuGlnThrAlaValAlaGluVal 900
|||||
2929 TGAAGAGAGTCAAGCTCATTTGAGAGAGTCTGCAGACCCAGTTGCCAGGTC 2978
901 ThrLeuAsnGlyAspIleProPheHIsArgProArgProArgIleIleArg 917
|||||
2979 ACTTTGAATGGGGATATCTCTTCCATCGCCACGCGCGGAGTAATACAG 3028
917 ggiValArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyrA 934
|||||
3029 AGGCAGGGATGCATAGAGAGTCTCAGAGTCCATCGACAAATCTTATG 3078
934 spLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950
|||||
3079 ATTAACCCCGCGTAGATGATGATGAAGGCGATGGAGACCTTGGAAGAA 3128
951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
|||||
3129 GACACAGAAAGTAAGTCAAGATCAGGAGAGAGTCTGTATCCACACAGCAG 3178
967 rAspGlnProSerHIsSerLeuGluArgGlnMetAsnGlyAsnGluGluL 984
|||||
3179 TGAATCAGCCTTCCCACTCTCTGAGAGACAAATGAATGGAAACCAAGAGA 3228
984 ysgGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
|||||
3229 AAGGTGATTAAGACTGATAGAAAAAAGATTAACCTGAAAAAAGAAAGAAC 3278
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
|||||
3279 AAAGATAGAGATTAAGAGAAAGATTAATAATGNAAGCAAGAAAGGAATGCT 3328
1017 ulLysGlyLeuGlyLysAspMetPheArgPheGlyLysHIsArgLysAspAspL 1034
|||||
3329 GAAGGGCTTTGGGACATGTTCAAGTTTGGCAAAACATCGAAAAAGATGACA 3378
1034 ysiIleGluLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050
|||||
3379 AGATTGAGAAAAAGGGTAAATAATACAGGAATCCTTTACATCAGAA 3428

1051 GluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysTh 1067
|||||
3429 GAGAGAGAGATACGAATGAAGCAGAGCAGAGAGAGATTCACCAAAAC 3478
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrAlaGluIleG 1084
|||||
3479 TCGAGAAATTTAGGAAAGACAGACAGCTCAGAGCGTGACTATGTAATTC 3528
1084 InAspPheHIsArgThrPheGlyCysAspAspGluLeuMetTyrGlyGly 1100
|||||
3529 AAGATTTCATCGACATTTGGCTGTGATGATGATTAATGATGGGGA 3578
1101 ValSerSerTyrGluGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
|||||
3579 GTTCTCTTATGAAGATTCTCATGGCTCTCAACGCTAGACCTCAGAGCC 3628
1117 oArgGluGlyHIsMetMetAspAlaLeuTyrAlaGlnValLysLysProA 1134
|||||
3629 ACAGAAAGGCGATATGATGATGCTTTGTATGCCAAAGTCAAGAAAGCCGC 3678
1134 rAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
|||||
3679 GGAATTCCAAAACCTCACCTGTAGACAGTAAACAGATCAACTCCTAGCAAT 3728
1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
|||||
3729 CATGATCGATACAGCGCTGTAGGCAAGAAATTCAGCCAAAGCAAGA 3778
1167 pGluAspValGluAspArgArgArgThrTyrSerPheGluGlnProTrp 1184
|||||
3779 TGAAGATGTAGAAAGATCTGCGCGAGACTTATAGCTTGTGACCAACCTCGCG 3828
1184 roAsnAlaArgProAlaThrGlnSerGlyArgHIsSerValSerValGlu 1200
|||||
3829 CGAAGCAGAGCGCGCGGAGCAGCAGAGCGGCGACACTGGTCTCGTGCAG 3878
1201 ValGlnMetGlnArgGlnArgGlnGluGluArgGluSerSerGlnGlnAl 1217
|||||
3879 GTGCAGATGCGAGCGCGCCAGCCAGAGAGAGCGGAGAGACTCCACAGCGGC 3928
1217 agInArgGlnTyrSerSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234
|||||
3929 CCAGCGCCACTACAGCTCTCTGCTCGGCAAGCAGGAAAAATGCCAGCT 3978
1234 eValSerGlnAspSerTyrPglGlnAsnTyrSerProGlyGluGlyPhe 1250
|||||
3979 CGGTCTCCAGAGACTCTTGGAGCAGAACTACTCCCTGGGGAAGGCTTC 4028
1251 GlnSerAlaLysGluAsnProArgTyrSerSerTyrGlnGlySerArgAs 1267
|||||
4029 CAGAGTGCCAAAAGAACCCAGGTACTCCAGCTACCAAGGCTCCAGAA 4078
1267 ngLytYrLeuGlyGlyHIsGlyPheAsnAlaArgValMetLeuGluThrG 1284
|||||
4079 CGGCTACTGTGAGAGACTGCTTCAACGCGCAGGCTCATCTGTGAAACTC 4128
1284 lngLeuLeuAspArgGlnGluGlnArgArgLysGlnGlnGlnMetLysLys 1300
|||||
4129 AGGAGGCTCTTCGCCAGAGAACAGAGGCGGAGAGAGCAGAGATGAAGAG 4178
1301 GlnProProSerGluGlyProSerAsnTyrAspSerTyrLysLysValG1 1317
|||||
4179 CACCTCTCTTCCGAGGGGCCAGCACTATAGACTCGATTAAGAAAGTCCA 4228
1317 naAspProSerTyrAlaProProLysGlyProPheArgGlnAspValAProP 1334
|||||
4229 GGAAGCCAGTTACGCCCTCCCAAGGGGCGCTTCCGGCAAGATGTGCCCC 4278
1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350
|||||
4279 CCGCCCTTCTCAGGTTGGAGGCTGAACAGACTCTGAGACTCTGAGAA 4328
1351 GlyArgProPheTyrSer 1356

|||||
4329 GGAGGCCCTTCTATCC 4346

seq_name: gb_pr:AF467004

seq_documentation_block:

LOCUS AF467004 3960 bp mRNA linear PRI 07-FEB-2002
DEFINITION Homo sapiens partitioning-defective 3 protein splice variant d
(PARD3) mRNA, complete cds, alternatively spliced.
ACCESSION AF467004
VERSION AF467004.1 GI:18568347

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 3960)

Gao, L., Macara, I.G. and Joberty, G.
Multiple splice variants of PAK3 and of a novel related gene,
PARD3, produce functionally different proteins

Unpublished
2 (bases 1 to 3960)

Gao, L., Macara, I.G. and Joberty, G.
Direct Submission

Submitted (10-JAN-2002) Center for Cell Signaling, University of
Virginia, PO Box 800577, Charlottesville, VA 22908, USA

Location/Qualifiers

1..3960

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..3960

/gene="PARD3"

1..3960

/note="PARD3"

/note="PARD3: multi-PDZ domain containing protein;
alternatively spliced"

/codon_start=1

/product="partitioning-defective 3 protein splice variant
d"

/protein_id="P467004.1"

/db_xref="GI:18568348"

/translation="MKTVCFGRTRVVPCGDMKVFSLIOAVRRYRAIKDPY
WIOVRLHGDGILDDILCDVADKRLVAFPDOPDHGDDGTSASITQSP
IFGSELTNNVSAFOPOATSETEVPSYRAMPLIHRSSDPALIGLSTVSDNF
SSEPSRKNTPTSTTGTGLKQNTAGSPKTCDDKRDENYSLRDRSNNMNOQRNA
RSLSASHPWGKMLEKQEDDEDEEDNSRVEPVADGLIHINFSIDVAKLVE
VPNDGGLIHVVPFSARGRTGLIVKLEKGAHEHLEFENCIVRINDGLRN
RRFOADHMFROAMRTPIIPFHVPAANKEOYRQLOSSEKNYSSRFPDSQYDNR
SVNSAGLHTVOAPRLNHPPEOIDSRLPHSAHPGKPPSAPASAPOVVESTVSSG
YNTKTIKRLNIOLKGTGEGITSGITSROVTIGGSAPYIKYKILPFGAIIODRLAG
DRLIEVGVYLVKSOEYVSLRSTKMEYSLVAFRODAHPRELNAEPQOQIP
KETKADDEDVLPDGTREFLITEVPLNDSGSLGVSVGNKSNHNDLGIYVSI
INGGASKDRLRVNDQLIAVNGESLIGTNDAMETLRSKSTEBKNGMLOLYAR
RISKNEMLSPSPGPBELPIETALDREIRLSHLSYSGIEGLDESPHNAALSRIMG
ESGKYLSPTVNMPODDTVIIEEDRLVLPPLHSDQSSSHDVGFFVADGTAKA
AIDSDADCSLSPVDVPLAFORREGFGROKSEKRTQSDASOLDPVRKRKSNMIG
IADETKLNTVDDOKAGSPSRDVGSGIKSSLESLQYVAEYVITNGIIPHRPRPR
IIRGRCNSFRAIDKSYDKRAVDDDDGEMETLEDTESSSGSESVTASDPSH
SLEROMANGOEKDKTDRKDKTGKKEKKDRDEKDKAKKGMKGLIDMFRIOAKT
REFRERKARBRDAETIODEHRTFGCDDELMTGVSSYEGSMALNAPSPRIGHMIDA
LYAQVKPRKSPKSPVDSNRSTPSNDRIORLROEQKQKODEVDRRRTVSPBPDW
PNARPAQSGRIHSVEYQOROROEERESSQAOQVYSILRQSKKNSVSGQSMW
ONYSPEGFOGSAKENPRYSYSGSRNGYTGGAQFNARVLEITQELRQDRKEQDMK
KOPSPSPSNYDSYKVVQPSYAFPRGPRRODVPSPSOVARLNRLQPEKGRPPYS"

814..1074

/gene="PARD3"

/note="Region: PDZ domain"

1378..1638

/gene="PARD3"

/note="Region: PDZ domain"

1765..2028

/gene="PARD3"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 1170 a 910 c 1048 g 832 t
ORIGIN

alignment_scores:

Quality: 6818.50 Length: 1336
Ratio: 5.169 Gaps: 1
Percent Similarity: 97.271 Percent Identity: 97.271

alignment_block:

US-09-757-781-2 x AF467004 ..

Align seg 1/1 to: AF467004 from: 1 to: 3960

1 MetLysValThrValCysPheGlyArgThrArgValAlaValProCysG1 17
1 ATGAAAGTACCGTGGTGGTGGACGACCGGGTGTCTGCTGCGCGG 50
17 yaspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
51 GGACGGCCACATGAAAGTTTCAGCCTCATCCAGCAGCGGCGACCCGCT 100
34 yfargLysAlaIleAlaLysAspProAsnTyrrPILeGlnValHisArg 50
101 ACCGGAAGGCCATCGCCCAAGATCCAAACTACTGATACAGTGCATCGC 150
51 LeuGluHisGlyAspGlyGlyIleLeuAspLeuAspPILeuCysAs 67
151 TTGGAACTGAGATGAGAGATACATGACCTGATGACATTCCTTGTGCA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
201 TGTAGCAGACGATTAAGACAGACTGTGAGCATGTGTGATGAGCAGATC 250
84 roHSHISGLYLYASPGLYTHRSERLASERTHGLYTHGLINER 100
251 CACATCAGCAGGAGTGTGAGCAGCAGTGCAGTTCACGGGTACCCAGAC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
301 CCAGAGATATTGTTAGTAGACCTTGGCACAACAATGTCACGCTTTC 350
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAACCAACAAGTGAATTTGAGTGCACACCTTCACTGCTTGAG 400
134 laasnmetProLeuHisValAlaArgArgSerSerAspProAlaLeuIleGly 150
401 CAAATATGCTCTTCATGATGTCAGAGCAGTAGTACCACCTTAATTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
451 CTCGCCACTTCTGTGACATGATGATTAATTTTCCCTGTAACAGCTTCAAG 500
167 gLysAsnProThrArgTrrPserThrAlaGlyPheIleuLysGlnAsnT 184
501 GAAAAATCCACACAGCGTGTGCAACAACAGCTGCTTCCCTCAACAGAAC 550
184 hrAlaGlySerProLysThrCysAspArgGlyLysAspGluAsnTyrrArg 200
551 CTGCTGGAGTCTCTAAACCTGCGACAGGAAGAAATGAAACTCTACAG 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGluPheGlnArgAsp 217
601 AGCCTCCCGGGGATACTACTAATCTGCTCAACCAATTTCCAGAGACAA 650
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
651 TGTCTGCTGCTCTCTGAGTGCAGTCAACCAATGTGGGCAAGTGTGCTGG 700
234 LuysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250

```
701 AGAACCAACAGATGAGATGGACAGAAAGATTAACAGTCTGTT 750
251 GIUProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhse 267
|||||
751 GAACCTGTTGACATGCTGACACGGGTTTGAGCATATATACCAACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
|||||
801 TCTGGATGATATGTGTAAGCTCTAGACGTCCTCAACGATGGAGGCGCTC 850
284 euGlyIleHisValIleProPhseSerAlaArgLysGlyArgThrLeuGly 300
|||||
851 TGGGAATTCATGTAGTGCCTTTTACGTCTCGAGCGCGCAGAACCTGGGG 900
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLe 317
|||||
901 TTATTTAGTAAACGATTTGGAGAAAGGTGAAGCTGAACATGAAATATCT 950
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
|||||
951 TTTTCTGTGAGAAATGATTCATGTCAGGATTAATGATGGCGACTTCGAA 1000
334 snArgArgPheGluGluAlaGluHisMetPheArgGluAlaMetArgThr 350
|||||
1001 ATAGGAAGATTGAAACAAGCAACATATGTTTCCCAAGCCATGCGTACA 1050
351 ProIleIleThrPheHisValValProAlaAlaAsnLysGluGlnArg 367
|||||
1051 CCCATCATTTGGTTCCATGTGGTTCCTGCGACGAATTAAGACAGATATGA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrylYrSerSerArgPheSer 384
|||||
1101 ACAACATCTCCAAAGTGAAGAAACAATTACTATTCAAAGCCGTTTAAACC 1150
384 roAspSerGlnTyrylIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
|||||
1151 CTGACAGCCCATATATTGACACACAGAGTGTGAACAGTGCAGGCGTTTCA 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProGluGlnIleAspSe 417
|||||
1201 ACGGTGACAGAGCACCCCGACTGAACACCCCGCTGAGCAGATAGACTC 1250
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSer 434
|||||
1251 TCACTCAAGCTACATATAGCCACACCCCTGGGAAACCAACATCCG 1300
434 laProAlaSerAlaProGluAsnValPheSerThrThrValSerSerGly 450
|||||
1301 CTCACAGCTCGGCACCTCAGAAATGTTATTAGTACGACTGTAAAGCATGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
|||||
1351 TTATTAACACCAAAAAAATAGCCAGAGGCTTAATATCCAGCTTAAGAAAG 1400
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
|||||
1401 TACAGAGAGGTTTGGATTCAGCATCACTTCACAGATGAACAATAGCTG 1450
484 LysSerAlaProIleTyrylValLysAsnIleLeuProArgGlyAlaAlaIle 500
|||||
1451 GCTCAGCTCCAAATCTATGTGAAAAACATTTCTCCCGGGGGGGCGCATTT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
|||||
1501 CAGGATTCGCGCAGCTTAAGGACAGAGACACTATATAGAGTAATGAGT 1550
517 laAspLeuValGlyLysSerGlnGluLysValSerLeuLeuArgSer 534
|||||
1551 AGATTAGTGGCAAAATCCCAAGAGGAAGTGTTCGCTTGGAAGCA 1600
534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
|||||
1601 CCAAGATGGAGGAACGTGAGGCTTCTGCTTTTCCGCAAGAAAGCGCC 1650
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
|||||
1651 TTCCACCAAGGGAACGTGAATGCAGAGCCAAAGCCAGATGCAGATTCCAA 1700
567 sGluThrLysAlaGluAspGluAspIleValIleLeuThrProAspGlyThra 584
|||||
1701 AGAAGCAAAAGCAGAAAGATGAGATATGTTGTTTACACCTGATGGCCCA 1750
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
|||||
1751 GGGAAATTCGACATTTGAAATCCCACTTAATGATTAGGATGTGCAAGCC 1800
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
|||||
1801 CTTCGTCTCAGTGTCAAAAGGTAAACCGGTCAAAAGAGAACACCGAGATT 1850
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAsp 634
|||||
1851 GGGAAATCTTGTCAAGTCCATTTATTAATGAGAGGACGATCTTAAGATG 1900
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
|||||
1901 GAAGGCTTCGGGTGAATGATCACTGATACGATTAATGGAGAAATCCCTG 1950
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
|||||
1951 TTGGGCAAGCAAAACCAAGATGCCATGGAACCTTAAGAGGCTTATGTC 2000
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
|||||
2001 TACTGAAGGCAATTAACGAGAAATGATCCAGCTTATGTTGCCAAGAGAA 2050
684 leSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700
|||||
2051 TAAACAAGTGCATGAGCTGAGTCAAGTCACTGGAGACCCCTTGACCTGAG 2100
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
|||||
2101 CTGCGCCATTGAACACGCGTTGATGATAGAGAAAGAAATTTCCCATTC 2150
717 rLeuTyrylSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
|||||
2151 CCTTCAAGTGGAGTTGAGGGGCTTGATGAATCCCGCAGCAAGAAATCTG 2200
734 laLeuSerArgIleMetGlyGluSerGlyLysTyrylGlnLeuSerProThr 750
|||||
2201 CCTCAGATGAGATATAGGGTGAAGTCAAGTAAATACCAAGCTGTCCCTACA 2250
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
|||||
2251 GTGAATATGCCCCCAAGATGACACTGTCATTATAGAAGATGACAGTTGCC 2300
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAsp 784
|||||
2301 AGTGCCTTCCTCCACATCTCTCTGACAGTCCCTCTTCCAGCTCCCATATG 2350
784 sPValGlyPheValThrAlaAspAlaGlyThrTPAlaLysAlaAlaIle 800
|||||
2351 ATGAGGGGTTTGTACCGGACGATGCTGTACTTGGCCAMGGCTGCATTC 2400
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
|||||
2401 AGTGATTACAGCCGACGCTGCTTTGAGTCCAGATGTGATGCAGTTCCTGC 2450
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGlyLysArgThrL 834
|||||
2451 TTTTCAACGAGAAAGATTTGACGTCAGAGTATCTCAAGAAAAAGCAACA 2500
834 yGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
|||||
2501 AGCAATTTTCAGATGCCAGTCAATTTGATTTTCGTTAAACAGCAAAATCA 2550
```

851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
 |||||
 2551 AAAACCATGGATTATAGTATAGCTACACAGACTAACTCAATACAGTGA 2600
 |||||
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 |||||
 2601 TGACCGAAGAACAGGTCTCCACGACAGAGATGTGGTCTTCCCTGGGTC 2650
 |||||
 884 eulLysSerSerSerLeuGlnLysLeuGlnThrAlaValAlaGluVal 900
 |||||
 2651 TGAAGAGTCAAGCTCTGGAGAGTTCAGACCGCAGTTGCCGAGGTG 2700
 |||||
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgIleIleArg 917
 |||||
 2701 ACTTTGATGGGATATCTCTTCCATCTCTCCACGGCGCGGATATATCAG 2750
 |||||
 917 gGlyArgGlyCysAsnGlnLysPheArgAlaAlaIleAspLysSerTyr 934
 |||||
 2751 AGCGAGGGGATGCAATGAGCTTCAGAGCTGCCATCGACCAATCTTATG 2800
 |||||
 934 sPLysProAlaValAspAspAspAspGluGlyMetGluThrLeuGlnL 950
 |||||
 2801 ATTAACCCGCGGTGATGATGATGATGATGATGATGATGATGATGATG 2850
 |||||
 951 AspThrGluLysSerArgSerArgSerGlyArgGluSerValSerThrAla 967
 |||||
 2851 GACACAGAAAGAGTTCAGATCAGGAGAGAGAGCTGTATCCACAGCCAG 2900
 |||||
 967 rAspGlnProSerHisSerLeuGlnLysArgGlnMetAsnGlyAsnGln 984
 |||||
 2901 TGATCAGGCTTCCCTCTCTGAGAGACAAATGATGATGATGATGATG 2950
 |||||
 984 yGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
 |||||
 2951 AAGGTGATAGACTGATAGAAAAAGATTAACGTGAAAAAGAAAGAAC 3000
 |||||
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
 |||||
 3001 AAAGATAGAGATAGAGAGAGATGATGATGATGATGATGATGATGATG 3050
 |||||
 1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAsp 1034
 |||||
 3051 GAAGGGCTTGGGAGACATCTTC..... 3072
 |||||
 1034 ySileGluLysThrGlyLysLysLysLysLysLysLysLysLysLys 1050
 |||||
 3072 3072
 |||||
 1051 GlnGluLysArgLysMetLysGlnGlnGlnGlnGlnGlnGlnGlnGln 1067
 |||||
 3073AGGATTCAGAGCCCAAAAC 3089
 |||||
 1067 rArgGluPheArgGlnArgGlnAlaArgGlnArgGlnArgGlnArgGln 1084
 |||||
 3090 TCGAAGATTATAGGAGACAGACAGCTCGAGCGCTGACTATGCTGAATTC 3139
 |||||
 1084 LnsAspPheHisArgThrPheGlyCysAspAspGlnLeuMetTyrGly 1100
 |||||
 3140 AAGATTTCATCGACATTTGGCTGTGATGATGATGATGATGATGATGATG 3189
 |||||
 1101 ValSerSerTyrGlnGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
 |||||
 3190 GTTCTCTCTATAGAGCTTCATGCTTCACGCTTCAGACCTCAGAGCCC 3239
 |||||
 1117 oArgGlnGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLysPro 1134
 |||||
 3240 ACGAGAGGGGATATGATGATGATGATGATGATGATGATGATGATGATG 3289
 |||||
 1134 rGlnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
 |||||
 3290 GGAATTCCAAACCCCTCAGCTGAGACAGTAAACAGATCAACCTCAGCAT 3339
 |||||
 1151 HisAspArgLysGlnArgLeuArgGlnGlnGlnGlnGlnGlnGlnGln 1167

|||||
 3340 CATGATCGGATACAGCTCTGAGCCAGAGAAATTTACAGCAAGCAAGCAGA 3389
 |||||
 1167 pGluAspValGlnAspArgArgArgArgArgArgArgArgArgArgArg 1184
 |||||
 3390 TGAGATGTAGAGATCGTGGCGGACCTATAGTTTGAACAACCTGGG 3439
 |||||
 1184 rGlnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGln 1200
 |||||
 3440 CGAACCCAGCGCGCGAGCGAGCCGCCACACACTCGTCCGTGAG 3489
 |||||
 1201 ValGlnMetGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1217
 |||||
 3490 GTTCAGATTCAGCGGACGCGAGGAGGAGCGCAGAGCTCCAGCAGGC 3539
 |||||
 1217 aGlnArgGlnTyrSerSerSerLeuProArgGlnSerArgLysAsnAla 1234
 |||||
 3540 CCAGCGCCAGTACAGCTCTCTGCTGGCMAAGCAGGAAAAATCCAGCT 3589
 |||||
 1234 eValSerGlnAspSerTrpGlnGlnAsnTyrSerProGlyGluGlyPhe 1250
 |||||
 3590 CGGCTCTCCAGGACTCTTGGAGCAGAACTACTCCCTGGGGAAGGCTTC 3639
 |||||
 1251 GlnSerAlaLysGlnAsnProArgTyrSerSerTyrGlnLysSerArgAs 1267
 |||||
 3640 CAGAGTCCCAAGAGAACCCAGAGTCTCAGCTACCAAGAGCTCCAGGAA 3689
 |||||
 1267 nGlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGlnThr 1284
 |||||
 3680 CGGCTACTCGGAGGAGCATGCTTCAACGCCAGGCGTATGAGGAAATCTC 3739
 |||||
 1284 LngLysLeuLysArgGlnGlnGlnArgArgArgArgArgArgArgArg 1300
 |||||
 3740 AGGAGCTCTCTCGCAGAGAACAGAGCGGAGAGCAGCAGCAGAGAGAG 3789
 |||||
 1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysArgValG 1317
 |||||
 3790 CAGCTCTCTCCAGAGGGGCCAGCACTATGACTCGTATAGAAAGTCCA 3839
 |||||
 1317 nAspProSerTyrAlaProProLysGlyProPheArgGlnAspValProp 1334
 |||||
 3840 GGAACCCCAATTACGCCCTCCCAAGGGGCCCTTCGGCAAGATGTCGCC 3889
 |||||
 1334 rGlnProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGlnLys 1350
 |||||
 3890 CCGCCCTCTCAGGTCGAGGCTGAAACAGACTTCACACTCTGAGAAA 3939
 |||||
 1351 GlyArgProPheTyrSer 1356
 |||||
 3940 GGGAGGCCCTCTTATTCC 3957
 |||||

seq_name: gb_AF467006

seq_documentation_block: 3933 bp mRNA linear PRI 07-FEB-2002

LOCUS AF467006

DEFINITION Homo sapiens partitioning-defective 3 protein splice variant f (PAR3) mRNA, complete cds, alternatively spliced.

ACCESSION AF467006

VERSION AF467006.1 GI:18568351

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1. (bases 1 to 3933) Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Gao, L., Macara, I.G. and Joberty, G.

TITLE Multiple splice variants of PAR3 and of a novel related gene, PAR3L, produce functionally different proteins

JOURNAL Unpublished

REFERENCE 2. (bases 1 to 3933) Gao, L., Macara, I.G. and Joberty, G.

AUTHORS Direct Submission

TITLE Submitted (10-JAN-2002) Center for Cell Signaling, University of

JOURNAL

FEATURES
 source Virginia, PO Box 800577, Charlottesville, VA 22908, USA
 Location/Qualifiers
 1..3933
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..3933
 /gene="PAR3"
 /note="PAR3"
 /gene="PAR3"
 /note="PAR3"
 /note="PAR3"
 /note="PAR3P; multi-PDZ domain containing protein;
 alternatively spliced"
 /codon_start=1
 /product="partitioning-defective 3 protein splice variant
 f"
 /protein_id="AAU76046.1"
 /db_xref="gi:18568352"
 /translation="MKYIVYCGRTRVYVPCGDGIMKVESLIQANTRRKAIAPKOPN
 WIOVRLHSDGGLDDLDLIDCDVADDDKDLVAVDEDDPHHGSGTSSSTGTSP
 IFGSLGTNNNSAFQYQATSEIVTPSVLRANMPLHYRRSDPALIGLSTVSDSNF
 SSEPSRKNPTFMSTAGFLKONTAGSPKTCRDKDENYRSILPDTSMNSNPFQDNA
 RSSLSASHPMVQKLEKQEDDEGTEDNSRVEPYGHADTGLEHLPNSLDMVLYE
 VPNDGPIGLIYVPPSARGRTIGLLVLRLEKGAHEHNLFRENDCTVRIINDGLRN
 RREFQAHMFROAMRTPIITWYVPAANKEDYEQLSQSEKNNYTSRSPSQYIDNR
 SVNSGLHTVQAPRLNHPPEQIDSHSRILPHSAHSGKPPSPAPASAPQNVFTVSSG
 YNTKIGRLNLQIKKTEGLGFSITSDVYIGGSAPYIKNLIIRGAILDGRLLKAG
 DRLLENGVDLVGKQSEEVNLSRSTKEGVSLRVEDQHPRELKAEDEDIYLT
 PDGREFLTFEVLNDGSAIGSVGKRSKKNADIGIFKSIINGGAAGKDGRL
 VNDOLIAVNGESLIGKTODAMETLRSMSTPCKRMQOLIVARIISCNLKSQGS
 PPGEPLPIETALDDERKISHLSYIGIEGLDSEPRNALSLIMKIQSLSPVNNFOD
 DIVIILEDRLVPLPHLSQSSSSHDDVGEYTAAGTWAKAIDSACSLSPPVD
 VLAQREGGRQIADETKLTNYDDQKASPSRDVGSGLKSSLSLESLQTLVAEYTL
 NGDIPFHRPRPIIRGRCNESFFRAIIDSYDKPAVDDDEGMETLEEDTEERSSGR
 ESDVSPADSPSHLEIRONGNQEKDITDRKKDQKKEKKRDKBKMKAQKGLK
 GLGDMFRGKHKHKKDKIKETGKIKIOESPTSEERILRMKQERQOAKTREREQAR
 ERDYAEIODEFHRTCCDEDELMATGVSYTEGSAALNARPOSPEGHMDALTYAQVKPR
 NSKSPVDNSNRSTPSNHDRIQLRQEQAKODEVEDRRYTSFEQPMNAPRAQVS
 GRHSYVEMQORQERERESSQAQROYSSLPRQSRKNSVSDSWQNSPGEGF
 QSAKENPRYSYQSGSRNGYLGGHFNARMLTEORLSRKRKEQMKKPPSEGS
 NYDSYKRYODPEYAPPKPFRQDVPSPSQVARNRLRQPEKGRFPYS"
 814..1074
 /gene="PAR3"
 /note="Region: PDZ domain"
 1378..1638
 /gene="PAR3"
 /note="Region: PDZ domain"
 1726..1989
 /gene="PAR3"
 /note="Region: PDZ domain"
 misc_feature
 BASE COUNT 1164 a 899 c 1049 g 821 t
 ORIGIN
 alignment_scores:
 Quality: 6754.00 Length: 1356
 Ratio: 5.156 Gaps: 3
 Percent Similarity: 96.608 Percent Identity: 96.608
 alignment_block:
 US-09-757-781-2 x AF467006 ..
 Align seg 1/1 to: AF467006 from: 1 to: 3933
 1 MetLysValThrValCysPheGlyArgThrArgValValValProCysG1 17
 |||||||
 1 ATGAAAGTGAACCGGTCTTCGAGACGACCGGGTGGTGGTGGTGGG 50
 |||||||
 17 yAspGlyHisMetLysValPheSerLeuIleGlnAlaValThrArgT 34
 |||||||
 51 GAGACGGCCACATGAAAGTTTTCACGCTCATTCACAGCGGTCACCGCT 100
 |||||||
 34 yTrpGlySerAlaIleAlaLysAspProAsnTyrrTrrIleGlnValHisArg 50
 |||||||

101 ACCGGAAGCCATCGCCACAGATCCAAACTACTGGATACAGGTGCAATCC 150
 51 LeuGlnHisGlyAspGlyGlyIleLeuAspLeuAspAspIleLeuCysAs 67
 |||||||
 151 TTGGAACTATGAGATGAGAGAAATCTACACTTGTATGATCATCTTTTGCA 200
 |||||||
 67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGlnGlnAsp 84
 |||||||
 201 TGTAGCAGACGATGAAGACACACTGTGTAGCAGTGTGTATGATGACGAGATC 250
 |||||||
 84 roHSHISGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
 |||||||
 251 CACATCACAGGAGTATGGACAGTGCACGTCTCACGCGGTAACCAAGAGC 300
 |||||||
 101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
 |||||||
 301 CCAGAGATATTTGGTAGTACCTTGGCCACCAATGCTCAGCCTTTTCA 350
 |||||||
 117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
 |||||||
 351 GCCTTACCAACCAACAGTGAATTTAGAGTCACACCTTCAGTCCTTCAG 400
 |||||||
 134 IaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
 |||||||
 401 CAAATATGCCCTTCTCATGTTCGACGCACTAGTGAACCCAGCTTAATTGGC 450
 |||||||
 151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerArg 167
 |||||||
 451 CTCCTCCACTTGTGTACAGTAGATTAATTTTCTCTGTAAGACCTTTCAAG 500
 |||||||
 167 GlyAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
 |||||||
 501 GAAAAATCCACACAGCTGTCACAAACAGCTGGCTTCCTCAAGCAGAAC 550
 |||||||
 184 hrAlaGlySerProLysThrCysAspArgGlyLysAspGlnAsnTyrrArg 200
 |||||||
 551 CTGTGGGAGTCTTAAACCTGCAGCAGAGAAAGATGAAGAAATCAACGA 600
 |||||||
 201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
 |||||||
 601 AGCCTCCCGCGGATCTAGTACTGGTCTTAACCACTTTCAGAGAGAAC 650
 |||||||
 217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
 |||||||
 651 TGTCTGCTGTCTGTGAGTGCACAGTCAACCAATGGTGGCAAGTGGCTGG 700
 |||||||
 234 LuLysGlnGlnAspGluAspGlyThrGlnGluGlnAspSerArgVal 250
 |||||||
 701 AGAAACAAAGACAGATGAGATGGACAGAAAGATTAACAGTGTGT 750
 |||||||
 251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
 |||||||
 751 GAACCTGTGGACATGCTGACACAGCGGTTTGGAGCATATACCAACTTTTC 800
 |||||||
 267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
 |||||||
 801 TCTGATGATATGTGTAAGCTCGTAGAAGTCCCAACATGAGGAGGCGCTC 850
 |||||||
 284 euGlyIleHisValValProPheSerAlaArgGlyGlyArgThrLeuGly 300
 |||||||
 851 TGGGAATCCATGTAGTGGCTTTCAGTGTGCGAGGGGAGAAACCTGTGGGG 900
 |||||||
 301 LeuLeuValLysArgLeuGluLysGlyLysIleLysAlaGlnHisGluAsnLe 317
 |||||||
 901 TTATTATTAGTAAACAGATTGGAGAAAGTGTGAAGCTGAACATGAAGAACT 950
 |||||||
 317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
 |||||||
 951 TTTTCGTGAGATGATGATTCATGTCAGAGATTAAATGATGGCACTTGC 1000
 |||||||
 334 snArgArgPheGlnGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
 |||||||
 1001 ATGAAAGATTTTGAACAGCAACACATATATGTTCGCCAAGCCATGGCTG 1050
 |||||||

351 ProIleIleThrPheHisValProAlaAlaAsnLysGluIleThrG 367
 |||||
 1051 CCCATCATTTGGTTCCATGTGGTCTCTGCAGCAATAAAGAGAGATATA 1100
 367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
 |||||
 1101 ACACCTATCCCAAGAGAGAGAACAAATTAATTCACGCCGTTTATGCC 1150
 384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGluLeuHis 400
 |||||
 1151 CTGCAGAGCCAGATATATTGACACAGAGAGTGTGACAGTCTCAGGGCTTAC 1200
 401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
 |||||
 1201 ACGGTGCAGAGAGACACCCCGACTGAACACCGCGCTGAGCAGATAGACTC 1250
 417 rHisSerArgLeuProHisSerHisSerSerGlyLysProProSerA 434
 |||||
 1251 TCACCTCAAGACTACTCATAGCCACACACCCCTCGGGAAACACCATCCG 1300
 434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
 |||||
 1301 CTCACGCCCTCGGCACCTCAGAAATGTAATTAGACACTGTAAAGCAGTGT 1350
 451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
 |||||
 1351 TATTAACACCAAAAATAAGGAGAGAGGCTTAATATCCAGCTTAAGAAAG 1400
 467 YThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
 |||||
 1401 TACAGAAAGTTGGGATTCAGACATCACCTCCACAGATGTAAACAATAGGTG 1450
 484 LySerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
 |||||
 1451 GCTACGCTCCCATCTATGTGAAAAACATTTCTCCCGGGGGCGGCATTT 1500
 501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517
 |||||
 1501 CAGATGGCCGACTTAAGGACGAGACAGACTTATAGAGGTAAATGAGAT 1550
 517 laSpleuValGlyLysSerGlnGluValValSerLeuLeuArgSerT 534
 |||||
 1551 AGATTTAGTGGGCAATCCCAAGAGCAAGTTGTTGCTTGCTTGAAGAACA 1600
 534 hrLysMetGluGlyThrValSerLeuValPheArgGlnGluAspAla 550
 |||||
 1601 CCAAGATGAGAGAACTGTAGCCCTTCTGCTTTCGCCAGGAAGACGCC 1650
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
 |||||
 1651 TTCACCCCAAGGGAAGCTG..... 1668
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
 |||||
 1669AAAGCAAGAAAGATGAGGATTTGTTCTTACACCTGATGGACACA 1711
 584 rGluIlePheLeuThrPheGluValProLeuAsnAspSerGlySerIleAgly 600
 |||||
 1712 GGGAAATTTCTGACATTTGAAGTCCCACTTAATGATTCAGAGATCTCAGGCC 1761
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
 |||||
 1762 CTTGGGTGTAAGTGTCAAAAGTAAACCGGTCAAAAGAAACCAACCGCATTT 1811
 617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
 |||||
 1812 GGGAAATCTTTGCAAGTCCATTTAATGAGAGACACCATCTAAAGATG 1861
 634 LyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 650
 |||||
 1862 GAAGCTTCGGGTGAATGATCATCACTGATAGCAGTAAATGGAATCCCTG 1911

651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
 |||||
 1912 TTGGGCAAGACAAACCAAGATGCCATGGAAACCTTAAGAAAGTCTATGTC 1961
 667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
 |||||
 1962 TACGAAGGCAATTAACAGAGAAATGATCCAGCTTATTTGTCGAAGAGCA 2011
 684 leSerLysCysAsnGlnLeuLysSerProGlySerProProGlyProGlu 700
 |||||
 2012 TAACCAAGTGCATGAGCTTAAGTACCTGTGGAGACCCCTGAGACCTGAG 2061
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
 |||||
 2062 CTGGCCATTGAACAGCGCTTGATGTATAGAACGAAAGATTTCCCATTC 2111
 717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
 |||||
 2112 CCCTCAAGTGGGATTTGAGGGCTTGATGATCCACACAGCAAAATGCTG 2161
 734 laLeuSerArgIleMetGlyLysSerGlyLysTyrGlnLeuSerProThr 750
 |||||
 2162 CCCTCAGTAGATATATG.....GTTAAATACCAAGCTGTCCCTTACA 2202
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
 |||||
 2203 GTCAATATGCCCCCAAGATGACACTGTCAATTATAGAAATGACAGGTTGCC 2252
 767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
 |||||
 2253 AGTGCTTCTCCACATCTCTGACACAGTCTTCCACGCTCCCAATGATG 2302
 784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaAlaIle 800
 |||||
 2303 ATGTGGGTGTTTGTGACGGCAGATGCTGTACTTGGGCCAAGGCTCATATC 2352
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
 |||||
 2353 AGTGAATTCAGCCGACTGCTTGTGAGTCCAGATGTGATCCAGTCTTGGC 2402
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
 |||||
 2403 TTTTCAACAGAGAGATTTGGACGTCAG..... 2430
 834 ySglnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
 |||||
 2430 2430
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
 |||||
 2431ATAGCTGAGAGACTTAACCTCAATACAGTGA 2462
 867 PASpGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 |||||
 2463 TGACCAGAAACCAAGTTCTCCACACAGACATGTGGGCTTCTCCCTGGGTC 2512
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
 |||||
 2513 TGAAGAAGTCAAGCTCTTGGAGAGCTTGCAGAACCGAGATGGCGAGGTT 2562
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleLea 917
 |||||
 2563 ACTTTGAATGGGATATTTCTTCCATCGTCCACGGCCCGGAGTATATCG 2612
 917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyrA 934
 |||||
 2613 AGGCAGGGCGATGCAATGAGAGCTTCAGAGCTGCCATCGCAAAATCTTATG 2662
 934 sPlySProAlaValAspAspAspGlyGlyMetGluThrLeuGluGlu 950
 |||||
 2663 ATAAACCCCGGGTATGATGATGATGAAGCATGAGACCTTGGAGAGAA 2712
 951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlase 967

```

|||||
2713 GACACGAGAAAGTTCCAGATCAGGAGAGAGTCTGTATCCACAGCCAG 2762
967 TASPGLNProSerHisSerLeuGluArgGlnMetLysGlnGlnGlu 984
|||||
2763 TGATACGCTTCCCTCTCTGTGAGAGACAAATGATGAAACCAAGAGA 2812
984 ysgLysApLysThrAspArgLysLysAspLysThrGlyLysGluLys 1000
|||||
2813 AAGGTGATAGACTGATAGAAAAAGGATAAACTGGAAAAAGAAAGAG 2862
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMet 1017
|||||
2863 AAAGATAGAGATAGGAGAAAGGATAAATGAAAGCCAAAGAAAGGAAAGCT 2912
1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAsp 1034
|||||
2913 GAAGGCTTGGGAGACATGTTCAAGCTTGGCAACATCGAAAGATGACA 2962
1034 ysIlleGluLysThrGlyLysIlleLysIlleGlnLysSerPheThrSerGlu 1050
|||||
2963 AGATTGAGAAACGGGTAAATATAATACAGGATCTTTACATCAGAA 3012
1051 GlnGluArgLysLeuArgMetLysGlnGlnGluArgLleGlnAlaLys 1067
|||||
3013 GAGAGAGAGTACGAAAGAAAGAGAGAGAGAGAGATTCAGGCAAAAC 3062
1067 rArgGluPheArgGluArgGlnAlaArgLysArgAspTyrAlaGlu 1084
|||||
3063 TCAGAGATTTAGGGAACGACAGACAGCTCGAGAGCTGACTATGCTGAATTC 3112
1084 LnsAspPheHisArgThrPheGlyCysAspAspGluLeuMetLysGly 1100
|||||
3113 AAGATTTCATCGAGACATTTGGCTGATGATGATTAAGTATAGGGGA 3162
1101 ValSerSerTyrGluLysSerMetAlaLeuAsnAlaArgProGlnSer 1117
|||||
3163 GTTCTCTTATGAAAGGTTCCATGGCTCTCAAGCTGACCTCAGACCC 3212
1117 cArgGluGlyHisMetMetAspAlaLeuTyrAlaGlnValLysLys 1134
|||||
3213 ACAGAAAGGCAATATGATGATGCTTGTGATGCCAAGTCAAGAACGCCG 3262
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
|||||
3263 GGAATTCCAAAACCTCCTCAGTACAGTAAACAGATCACTCCATCAACA 3312
1151 HisAspArgLleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGln 1167
|||||
3313 CATGATGGATACAGCGCTGAGGCAAGAAATTCAGCAAGCAAGCAAGA 3362
1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProTrp 1184
|||||
3363 TGAAGATGTAGAGATGTCGGCGGACCTATAGTTTGACCAACCCCGGC 3412
1184 rAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
|||||
3413 CGAAGCCAGCGCGCGGACGACAGCGCGGACCTCGGTGTCGGTAGAG 3462
1201 ValGlnMetGlnArgGlnArgGlnGluLysGlnLysSerGlnGln 1217
|||||
3463 GTGAGATGACAGCGGCGGACGAGAGAGCGGACGAGCTCCAGCAGGC 3512
1217 agLnsArgLysSerSerLeuProArgLnsArgLysAsnAlaSer 1234
|||||
3513 CGAGCGCATACAGCTCTCTGCTCGCAAAAGAGAAAAATGCCAGCT 3562
1234 eValSerGlnAspSerTrpGluGlnAsnTyrSerProGlyGluGlyPhe 1250
|||||
3563 CGGCTCTCCAGAGACTCTTGGGAGCAGAACTACTCCCTGGGGAAGGCTTC 3612
1251 GlnSerLysGluAsnProArgTyrSerSerTyrGlnGlySerArgAs 1267
|||||

```

```

3613 CAGAGTCCCAAAAGAAACCCAGGTACTCCAGCTACCAAGGCTCCAGAA 3662
1267 nGlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGlnThrG 1284
|||||
3663 CGGCTACTTGGGAGACATGCTTCAACGCCAGGGTATGCTGAAACTC 3712
1284 lngLeuLeuArgGlnGlnArgArgLysGlnGlnMetLysLys 1300
|||||
3713 AGGAGCTCTTCCGCGAGAAACAGAGCGGAGGAGAGAGATGAAAGAG 3762
1301 GlnProProSerGluGlyProSerAsnTyrAspSerTyrLysValGlu 1317
|||||
3763 CAGCTCTCTCCGAGGGGCCAGCACTATGACTCGTATAAGAAAGTCCA 3812
1317 nasPProSerTyrAlaProProlysGlyProPheArgGlnAspValProp 1334
|||||
3813 GGACCCAGATTACGCCCTCCCAAGGGGCCCTTCCGGCAAGATGTGCC 3862
1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350
|||||
3863 CTTCCCTTCTCAGAGTTCGAGGCTGAACAGACTTCAGACTCTGAGAAA 3912
1351 GlyArgProPheTyrSer 1356
|||||
3913 GGGAGGCCCTTCTATTC 3930
seq_name: gb_pr:AF332593
seq_documentation_block: 3822 bp mRNA linear PRI 16-NOV-2001
LOCUS AF332593
DEFINITION Homo sapiens atypical PKC isotype-specific interacting protein long
variant b mRNA, complete cds, alternatively spliced.
ACCESSION AF332593
VERSION AF332593.1 GI:14579310
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3822)
Fang,C.M. and Xu,Y.H.
Down-regulated expression of atypical PKC-binding domain deleted
asip isoforms in human hepatocellular carcinomas
Cell Res. 11 (3), 223-229 (2001)
JOURNAL
MEDLINE 21520080
PUBMED 11642408
REFERENCE
2 (bases 1 to 3822)
Fang,C. and Xu,Y.
Exon/Intron Structure And Splicing Variants of A Novel Human
Polarity Gene, hasIP
JOURNAL
Unpublished
AUTHORS 3 (bases 1 to 3822)
Fang,C. and Xu,Y.
Direct Submission
Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
Oncology, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
320 Yue-Yang Road, Shanghai 200031, P. R. China
Location/Qualifiers
FEATURES
source
1..3822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..3822
/Note="ASIPb; alternatively spliced; similar to GenBank
Accession Number AF196185"
/codon_start=1
/product="atypical PKC isotype-specific interacting
protein long variant b"
/protein_id="AAK69193.1"
/db_xref="GI:14579311"
/translation="MKVTYCFGRTRVYVPCGSGIMKRVESLIQDAVTRKALAKDPY

```

WIOVHRLHEDGGLDLDLDLIDVADDKRLVAVFEDQPHHGSGTSASSTGTOSPE
IFGSELGTNNVSAFOQTOSTSEIYTPSVLRANMPLEHRRSSDALIGLSTVSDSNF
SSESRKNPTNMTSTTAGFLKONTAGSPKTDCKDKDENYRSLPDTNMSNOMORDVA
RSLLSASHMPVGMLEKQEQDDEGTEDNSKREYKADHDEHLPNLSLDDMKLVE
VPDGPPLGIHVVPESARGKRLGLLVKLEGGABEHENI.FRENDICVIR.IINDGLLN
RREFOAHMRQAMRPIIMFHVPAANKKEQJESQSEKNKYSSRPSPOXYIDNR
SYNSAGLHTVORAPRLNHPPEOIDSLSRLPHSAHSGKPSAPASAPONVSTVSSG
YNTKIKGRLNIOIKKGTGLGFSITSRDVTTGGSAPIYKVKLIFRGAIDODGLKKG
DLILEVNGDLVKSQSEYVSLRSTKMEGTVALVFRQEDAFHRELKADDEDIVLT
PDGTRFELTEVEPLNDSGSAICGVSKNRSKENHADLGI.FVKSIIINGAASDKRLK
VNDLI.VANGESILGKTNQDAMETLRPSTSEKNKGMQOLIVARIISKCNELKSPGS
PVGDELLETALDDREKRISHSLSGIBELDESPSRNALSLMGKYOOLSPTVNNPQD
DVTIIEDDRPLVPLPHLSQSSSSSHDDVGFYTPADGAKAKAISDASCSLSPVVD
VIAFORBEGFOIADDET.KLNTVDDQAKAGSPSRVAGSLGKSSSLEIOTIVAVETL
NGDIPFHHPRPRIIRGRGKNBSFPAIDKSYDPVADDDEGHETLEPTEERSRSGR
ESVSTASPOSHSLEROMNGNEKEDTDKDKTKEKKKDDKDKDKDMKAKKGMK
GLGDMFRLOAKTRRERROARERDYAEIODFHTFGCDELMATGSGSYEGSMALNAR
POSREGHMDALTAQVKKPNSKSPVDSNHRISLHRIORLEFOQAKODEDVE
DRRTYSFEOQWPMPNAPATOGSRHSVEVOMOROEERESSQAQORQYSLPQSR
KNASVSODSWEONTSPEEGFOSKAKENRYSYOGSNGYLGHGHNARVMELOBL
ROEORRKEQOMKOPPESEGPSNYSYKRVODPSIAPKGFQDVPSPSOVARLNRL
QPEKGRPEFS*

BASE COUNT 1114 a 885 c 1019 g 804 t
ORIGIN

alignment_scores: Length: 1356
Quality: 6537.50 Gaps: 4
Ratio: 5.136
Percent Similarity: 93.879 Percent Identity: 93.879

alignment_block:
US-09-757-781-2 x AF332593 ..

Align seg 1/1 to: AF332593 from: 1 to: 3822

1 MetLeuValThreValCysPheGlyArgThrArgValValProCysG1 17
|||||
1 ATGAAAGTACGCGTGTCTCGACGACCGCGGTGCTGCTGCGCGG 50
17 yAspGlyHisMetLeuValPheSerLeuIleGlnIleAlaValThrArgT 34
|||||
51 GGACGGCCACATGAAAGTTTCCGCTCATCCAGCGGCGTGACCCGCT 100
34 YrATGlySAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
101 ACCGGAAGGCCATGCCAAGATCCAAACTACGTGATACAGGTGCTCCG 150
51 LeuGlnHisGlyAspGlyIleLeuAspLeuAspAspIleLeuCysAs 67
|||||
151 TTGGAACATGGAGATGGAGAACTAGACCTTGATGACATCTCTTCTGA 200
67 pValAlaAspAspIlyAspArgLeuValAlaValPheAspGluGlnAsp 84
|||||
201 TGTAGACAGCATAAAGACAGCTGTGAGCAGTGTGATGACAGAGCAT 250
301 CCAGAGATATTTGCTAGCTGGCACCAACAATGCTCTGACGCTTTC 350
117 nProGlyGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
|||||
351 GCCTTACCAAGCAACAATGAGGTACACACTTCAAGTCCCTTCGAG 400
134 laaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
|||||
401 CAATATATCCCTTCATGTTGAGCGAGTAGTACCCAGCTCTCAANTTGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167

|||||
451 CTCGCCACTTCTGTGAGTGTAGTAATTTTCTCTCAAGACCCCTCAAG 500
167 gLyAspProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnTr 184
|||||
501 GAAAAATCCCAACAGCGTGGTCAACAACAGTGGCTTCTCTCAAGCAGMA 550
184 hTrAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTGCGGGAGTCTTAAACCTGCGACAGAAAGATGAAACTCAACA 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAsp 217
601 ACCCTCCCGGGGATACATGATACCTGCTTAAACCAATTTGAGAGACAA 650
217 naLaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
|||||
651 TGCTGCTGTCTGTGATGATGCCAGTCAACCCAAATGTTGGCAAGTGGCTG 700
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
701 AGAAACAAGACAGATGAGATGGGACAGAGAGGATTAACAGTCTGCTT 750
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
751 GAACCTGTGGACATGCTGACACAGGCTTTGGAGCATATACCACTTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284
801 TCTGATGATATGATGATTAAGCTCTGAAGATCCCAACGATGAGAGGCTC 850
284 euGlyIleHisValValProPheSerAlaArgGlyLysArgThrLeuGly 300
851 TGGGATCCATGTAGTGCCTTTCAGTCTGACGCGCGAGAACCTCGGGG 900
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGlnHisGlnAsnLe 317
901 TTATTAGTAAACGATTTGGAGAAAGGTGTAAGCTGAACCTGAACATGA 950
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
|||||
951 TTTTCTGAGAAATGATTTGATTTGACGATTAATGATGAGGACCTTGAA 1000
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1001 ATGAGAGATTGAAACAGCAACATATGTTTCCCAAGCCCATGCGTACA 1050
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrG 367
1051 CCCATCTTTTGGTTCATGTGGTCCGCGCAACCAATTAAGACAGTATGA 1100
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSer 384
1101 ACAACTATCCCAAGTGAAGAAACAAATTACATTCATCAACCGCTTTAGC 1150
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerIleGlyLeuHis 400
1151 CTGACAGCCAGTATATGACACAGAGAGTGTAACAGTGAAGGGCTTCA 1200
401 ThrValGlnAlaGluProArgLeuAsnHisProProGluGlnIleAspSe 417
1201 ACGGTGCGAGAGACACCCGACTGAACCAACCCGCTGACAGATAGATCT 1250
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSer 434
1251 TCACCTCAAGACTACTCTATACGCAACACCCCTCGGAAACCAACATCC 1300
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1301 CTCACACCTCGGACCTCGAATGATTTTATAGACAGTGAACAGTGGCT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467

1351 TATTAACCAAAAAATAGGCAAGAGCTTAATATCCAGCTTAAGAAAG 1400
 467 yThrgluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGly 484
 1401 TACAGAAAGGTTGGATTGCATCTCAGCTTCCAGAGATGTAAACAATAGGTG 1450
 484 ySerAlaProIleTyValValAsnIleLeuProArgGlyAlaIle 500
 1451 GCTCAGCTCCAAATCTATGTGAAAAACATTCCTCCCGGGGGGGCGCATT 1500
 501 GlnAspGlyArgLeuValAspArgLeuIleGluValAsnGlyVal 517
 1501 CAGAGTGGCCGAGCTTAAGCAGCAGACAGACTTATAGAGGTAAATGGAGT 1550
 517 LaspLeuValGlyLysSerGlnGluValValSerLeuLeuArgSer 534
 1551 AGATTATAGTGGCAAAATCCCAAGAGAGAGTTGTTGCTGTGAGAAAGCA 1600
 534 hTrpMetGluGlyThrValSerLeuValIleArgGlnGluAspAla 550
 1601 CCAAGATGGAGAGAGCTGAGCCCTTCTGCTTTCGCCAGGAAGAGCGCC 1650
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProly 567
 1651 TTCACCCCAAGGGAGCTG..... 1668
 567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
 1669AAAGCAGAAAGATGAGATATTGTTTCTTACCGTCGATGGCA 1711
 584 rGluLheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
 1712 GGGAAATTTGACATTGTGAAGTCCCACTTATATGATTCAGATCTGCAGGC 1761
 601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
 1762 CTGGGTGTCAGTGTCAAGGTAAACCGGTCAAAAGAGACACCGACAGTTT 1811
 617 uGlyIlePheValLysSerIleLeasnGlyAlaAlaSerLysAspG 634
 1812 GGGAAATCTTGTCAAGTCCATATTATATGAGAGCAGCATCTTAAAGATG 1861
 634 LyrArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
 1862 GAAGGCTTCGGGTGAATGATCACTGATACAGTAATGGAATGCCCTG 1911
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
 1912 TTGGGCAAGCAACCAAGATGCCATGGAAACCTTAAAGAGGTCTATGTC 1961
 667 rThrgluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
 1962 TACTGAAGCAATTAACGAGAGATGATCCAGCTTATGTTGCCAAGAGAA 2011
 684 IeSerLysCysAsnGluLeuLysSerProGlySerProGlyProGlu 700
 2012 TAAGCAAGTGCATGAGCTCACTGAGCCCTGGAGCCCTGGAGCTGAG 2061
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
 2062 CTGCCATTGAAACACAGCGTTGGATGATAGGAAGAGAAATTTCCCATTC 2111
 717 rLeuTySerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
 2112 CCTCTACAGTGGGATGAGGGGCTTGATGAATGCCCAAGCAAAATGCTG 2161
 734 lAlaLeuSerArgIleMetGlyGluSerGlyLysTyrglnLeuSerProThr 750
 2162 CCCTCAGTAGAATAAG.....GGTAAATAACAGAGCTGCCCTAC 2202
 751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
 2203 GTGAATATGCCCAAGATGACACTGCTATTATAGAGATGACAGGTTGCC 2252

767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
 2253 AGTCTCTCTCAACATCTCTCTGACACAGTCTCTTCCAGCTCCCATGATG 2302
 784 sPValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
 2303 ATGTGGGGTTTGTACGCGCAGATGCTGTACTTGGGCGCAAGCGTGCATC 2352
 801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
 2353 AGTCATTACGCCGCTCTCTCTTGTGATCCAGATTTGATCCAGTTCCTTCC 2402
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGlyLysArgThrL 834
 2403 TTTTCAACGAGAAGAGATTGGAGCTCAG..... 2430
 834 ySGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
 2430 2430
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
 2431ATAGCTGACGAGACTTAACCTAATACAGTGA 2462
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 2463 TGACACGAAGAGCAGCTTCTCCACAGAGATGTGGGTCTTCCCTGGGCTC 2512
 884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
 2513 TGAAGAGTCAAGCTCATTTGAGAGTCTGCAGACCGCAGTTCGCGAGCTG 2562
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917
 2563 ACTTTGAATGGGATATTCTTTCATCTGTCACGGCCGCGGATTAATCAG 2612
 917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyra 934
 2613 AGCGAGGGGATGCAATGAGAGCTTCAGAGCTGCATCGACAAATCTTATG 2662
 934 sPlysProAlaValAspAspAspAspGluGlyMetGluThrLeuGluGlu 950
 2663 ATTAACCCCGCGTATGATGATGATGAAAGCATGGAACCTTGGAAGA 2712
 951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
 2713 GACACAGAGAAATTTCAAGATCAGGGAGAGAGTCTGTATCCAGAGCCAG 2762
 967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnGlu 984
 2763 TGATCAGGCTTCCCACTCTCTGGAGAGCAAAATGAATGGAACCAAGAGA 2812
 984 ySGlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
 2813 AAGGTGATTAACACATGATAGAAAAAGATTAACCTGGAAAAAGAAAGAG 2862
 1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
 2863 AAACATAGAGATTAAAGAGATTAATAATGAACCAAGAGAGGAAATGCTT 2912
 2913 GAAGGCTTGGGAGACATGTTCC..... 2934
 1034 ySlleGluLysThrGlyLysIleLysIleGlnGluSerPheThrSerGlu 1050
 2934 2934
 1051 GluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysTh 1067
 2935AGGATTCAGGCCAANAAC 2951


```

misc.feature      /note="Region: PDZ domain 2"
                  1594..1851
                  /gene="PAR3"
BASE COUNT      1121 a 867 c 1016 g 797 t
ORIGIN

alignment_scores:
    Quality: 6485.00      Length: 1356
    Ratio: 5.122          Gaps: 4
    Percent Similarity: 93.363      Percent Identity: 93.363

alignment_block:
US-09-757-781-2 x AF252293 ..

Align seg 1/1 to: AF252293 from: 1 to: 3801

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
|||||
1 ATGAAAGTGCACCGTGCCTTCGAGACGACCGGGTGTGCTGCCGCGCG 50
17 YASPGLYHISMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
|||||
51 GGAGGCGCACATGAAAGTTTTCAGCCCTCATCCAGACGGCGTACCCGCT 100
34 YrArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
|||||
101 ACCGGAAGGCCATCGCCAAAGATCCAAACTACGTGATACAGTGCCATCGC 150
51 LeuGluHisGlyAspGlyIleLeuAspLeuAspAspIleLeuCysAs 67
|||||
151 TTGGAACATGAGATGAGCAATACATCTGACCTTGATGACATCTCTTGTA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
|||||
201 TGTAGCAGAGATTAAGACAGACTGGTAGCAGTGTGATGAGCAGGATC 250
84 rHisHisGlyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
|||||
251 CACATCACGAGGTGATGGCACACAGTCCAGTCCACGGGTACCACAGC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
|||||
301 CCAGAGATATTGTGTAGTGTGACCCACCAACAATGTCTCAGCCCTTCA 350
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
|||||
351 GCCTTACCAAGACAAAGTGAATGTAGGTCCACACCTTCAGTCTTGAG 400
134 LaaSerMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
|||||
401 CAATATAGCCTCTTCANGTTCGACGCGAGTAGTGACCCAGCTCTAATTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGlnProSerAr 167
451 CTCTCCACTCTCTGACGTGATTAATTTTCTCTGAAGAGCCCTTCAAG 500
167 gLysAsnProThrArgTrpSerThrThrAlaGlyPheLeuLysGlnAsnT 184
501 GAAAAATCCCAACGCGTGTCAACACAGCTGGCTCTCAAGCAGACACA 550
184 hAlaArgSerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTCTGGGATCTCCTAAACCTCGACAGCAAG..... 582
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArcGAspAs 217
582 ..... 582
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeug 234
582 ..... 582

```

```

234 LuLysGlnGlnGluAspGluAspGlyThrGlnGluAspAsnSerArgVal 250
|||||
583 .....GATGAGAGATGGGACACAGAAAGCATTAACAGTCGTGT 618
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPheSe 267
|||||
619 GAACCTGTGTGACATGCTGACACGGGTTTGGAGCATATACCAACTTTC 668
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284
669 TCTGATGATATGATTAAGCTCGTAGAATGCCCAACAGATGAGAGGCCCTC 718
284 euGlyIleHisValValProPheSerAlaArgGlyGlyArgThrLeuGly 300
719 TGGGAATCCATGTAGTCCCTTTCAGTCTCTGAGGCGGACAGACCTTGCG 768
301 LeuLeuValLysArgLeuGluLysGlyGlyLysAlaGluHisGluAsnLe 317
769 TTATTACTAAACGATTGGAGAAAGGTGTAAAGCTGAACATGAAAAATCT 818
317 upNeArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
819 TTTTCGTGAGATGATGTGCATGTGACAGATTAATGATGGCAGCTTGGA 868
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
869 ATGAAAGATTGGAACACAGCAACATATGTTGCCCAAGCATCGCTGACA 918
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGlnTyrG1 367
919 CCCATCATTTGGTTCATGTGTCTCCGACGAAATTAAGAGCGATGAGA 968
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
969 ACAACTATCCCAAGATGAGAAACAATTACTATTAAGCCGTTTATGCC 1018
384 roAspSerGlnTyrIleAspAsnArgSerValLaaSerAlaGlyLeuHis 400
1019 CTGACAGCCAGTATATTGACAAACAGAGTGTGAACAGTGCAGGCTTCAC 1068
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1069 ACGGTGCAGAGAGCACCCTGACTCAACACCCTGAGCAGATGATGACTC 1118
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1119 TCACCTCAAGACTACCTCATAGCGCACACCCCTCGGAAACCAACCATCCG 1168
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1169 CTCACAGCCTCGGCACCTCAGAAATGATATTAGTACAGCTGTAAGCAGTGT 1218
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG1 467
1219 TATTAACACCAAAAAAATAGCCAAAGAGCTTAATATCCAGCTTAAGAAAG 1268
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1269 TACAGAAAGTTTGGATTCAACATCACTCCAGAGATGTAACAATATAGTGTG 1318
484 lSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1319 GCTAGCTCCAAATATATGTGAAAAAANTTCCTCCCGGGGGGGCCCATTT 1368
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValLaaGlyVa 517
1369 CAGATGCGCGACTTAAGCGCAGGACAGACTTATAGAGTAAATGGAGT 1418
517 lAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
1419 AGATTTAGTGGCAAAATCCCAAGAGAGAGTGTTCCTGCTGTGAGAAACA 1468

```

534 hrlysmetClnGlyThrValSerLeuValPheArgInGlnuAspAla 550
 |||
 1469 CCAAGATTGGAGAACTGTGAGCCTTCTGTCTTCCGACGAGAAACGCC 1518
 551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
 |||
 1519 TTCACACCCAAAGGAACCTG..... 1536
 567 sgLutThrLyAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
 |||
 1537AAAGCAGAAAGATGAGATATCTTCTTACACCTGATGGCACA 1579
 584 rgluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
 |||
 1580 GGAATTTTCGACATTTGAAGTCCCACTTAATGATTCAGGATCTGCAGCG 1629
 601 LeuGlyValSerVallysgLyAsnArgSerLyGluAsnHisAlaAspLe 617
 |||
 1630 CTGGGTGTCAAGTCAAAAGGTAAACCGTCAAAAGACACCCACACATT 1679
 617 uGlyIlePheVallysserIleIleAsnGlyGlyAlaAlaSerLyAspG 634
 |||
 1680 GGGATCTTTGTCAAGTCCATTTATATGAGAGAGCAGCATCTAAAGATG 1729
 634 LyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
 |||
 1730 GAAAGCTTCGGGTGAATGATCACTGATAGCAGTAAATGGAGAAATCCCT 1779
 651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgSerMetSe 667
 |||
 1780 TTGGGCAGAACCAACCAAGATGCCATGGAAACCTAAGAAAGTCTATGTC 1829
 667 rThrGluGlyAsnLyArgGlyMetIleGlnLeuIleValAlaArgArg 684
 |||
 1830 TACGAAAGCAATTAACGAGGAATGATCCAGCTTATTTGTCAGAGGAA 1879
 684 IeserLyCyAsnGlnLeuLysSerProGlySerProGlyProGlu 700
 |||
 1880 TAAAGAACTGCATGAGCTGAACTGACCTGGAGACCCCTGGAGCCGAG 1929
 701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHis 717
 |||
 1930 CTGCCCATTTGAACAGCGTTGATGATAGAGAACGAAATTTCCCATTC 1979
 717 rLeuTySerGlyIleGlyGlyLeuAspGluSerProSerArgAsnAla 734
 |||
 1980 CCTCTACAGTGGGATTGAGGGCTTGATGAATCCCAAGCAGAAATGCTG 2029
 734 lAlaSerArgIleMetGlyGluSerGlyLysTyrgInLeuSerProThr 750
 |||
 2030 CCTCAGTAGATATATG.....GGTAAATACCACTGTCCCTCACA 2070
 751 ValAsnMetProGlnAspAspThrValIleIleGlnuAspAspArgLeuPr 767
 |||
 2071 GTGAATATGCCCAAGATGACACTGTCAATTAAGAAATGACAGGTTC 2120
 767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
 |||
 2121 AGTGTCTCTCCACATCTCTGACACAGTCTCTCCAGCTCCCATGATG 2170
 784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
 |||
 2171 ATGTGGGGTTTGTGACGCGCAGATGCTGTACTGGGCGCAAGGCTGCATC 2220
 801 SerAspSerAlaAspCySerSerLeuSerProAspValAspProValLeuAl 817
 |||
 2221 AGTGATTCACCGACTGCTCTTTGAGTCCAGATGTGATCCAGTCTTCTGC 2270
 817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
 |||
 2271 TTTTCAACGAGAAAGATTGGAGCTCAG..... 2298
 834 ysgInPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850

2298 2298
 851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
 |||
 2299ATAGCTGACGAGACCTTAACTCAATACATGTGA 2330
 867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
 |||
 2331 TGACCAAGAAAGACAGTCTCCACAGAGATGTGGGTCTCTCCCTGGGTC 2380
 884 euLyLysSerSerSerLeuGlySerLeuGlnThrAlaValAlaGluVal 900
 |||
 2381 TGAAGAAGTCAAGCTCGTTGGAGAGTCTCCAGACCGCAGTTGCCGAGGTG 2430
 901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleLea 917
 |||
 2431 ACTTGAATGGGATATTCCTTTCATCTGTCACAGCGCGCGGAAATACAG 2480
 917 gGlyArgGlyCyAsnGlnuSerPheArgAlaAlaIleAspLysSerTyra 934
 |||
 2481 AGCGAGGGGATGCAATGAGACTTCAGAGCTGCCATCGACAAATCTTATG 2530
 934 sPLysProAlaValAspAspAspGluGlyMetGluThrLeuGluGlu 950
 |||
 2531 ATMAACCCGCGGTAGATGATGATGATGAGGCAATGATGATGATGATGAT 2580
 951 AspThrGluLysSerSerArgSerGlyArgGlySerValSerThrAlaSe 967
 |||
 2581 GACACAGAAAGAAAGTTCAAGATCAAGGAGAGAGCTGTATCCACAGCCAG 2630
 967 rAspGlnProSerHisSerLeuGluArgGlnMetLysnGlnGlnGlu 984
 |||
 2631 TGATCAGCTCTCCACTCTCTGTGAGAGACAAATGATGATGAAACCAAGGA 2680
 984 yGlyLysPlyThrAspArgLysLysAspLysThrGlyLysGlyLysLys 1000
 |||
 2681 AAGGTGATAGACTGATGACAAAGAAAGAAATCGAAAGAAAGAAAGAAAG 2730
 1001 LysAspArgAspLysGlyLysAspLysMetLysAlaLysLysGlyMetLe 1017
 |||
 2731 AAAGATAGAGATTAAGGAGAAAGATTAAGAAAGCAAGCAAGGAAATGCT 2780
 1017 uLysgLyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034
 |||
 2781 GAAAGGCTTGGAGACATGTTCAAGTGTGGCAACATCGAAAGAAATGACA 2830
 1034 yslIleGluLysThrGlyLysIleLysIleGlnGlnuSerPheThrSerGlu 1050
 |||
 2831 AGATTGAGAAACGGGTAAATTAATTAATACAGGAATCCTTTACATCAGAA 2880
 1051 GlnGluArgIleArgMetLysGlnGlnGlnGluArgIleGlnAlaLysTh 1067
 |||
 2881 GAGAGAGAGATGAGATGAGAACGACGACGAGAGAGAGATTCAGACCAAAAC 2930
 1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyraGluIleG 1084
 |||
 2931 TCGAGAAATTAAGGAAACGACACAGCTGAGAGCTGATGCTGAAATTC 2980
 1084 lAspPheHisArgThrPheGlyCyAspAspGluLeuMetTyrglyGly 1100
 |||
 2981 AAGATTTTCAATCGGACATTTGGCTGTGATGATGATGATGATGATGATG 3030
 1101 ValSerSerTyrgLysSerMetAlaLeuAsnAlaArgProGlnSerPr 1117
 |||
 3031 GTTCTCTTATGAAGTTCATGCGCTCAACGCTAGACCTCAGAGGCC 3080
 1117 oArgGluGlyHisMetAspAlaLeuTyraGlnValLysLysPro 1134
 |||
 3081 ACGAAGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 3130
 1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
 |||

34 yTArgLysAlaIleAlaLysAspProAsnTyrTrpIleGlnValHisArg 50
101 ACCGAAAGCCATCGCCAAAGATCCAAACTACTGATGACGGTGATCCG 150
51 LeuGlnHisGlyAspGlyGlyIleLeuAspLeuAspAspIleucysAs 67
151 TTGGAAACATGAGATGAGAAATATACACTTGATGACATCTTTGTGA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
201 TGTAGACAGACGATTAAGACAGACTGCTAGCAGTGTGATGACGACGATC 250
84 roHisHisGlyLysGlyLysSerAlaSerSerThrGlyThrGlnSer 100
251 CACATCACGAGAGGTGATGGCACACAGTCCAGGTTCCACGGTACCCAGAG 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerIlePheG 117
301 CCAGAGATATTTGGTAGTGAGCTTGACACCAACATGTCTCAGCCCTTTCA 350
117 nProTyrGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAAGCACACAGTGAATTTAGGTGACACCTTCAGTCTTCGAG 400
134 IAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
401 CAAATATGCTCTTCATGTTCGACAGTAGTACCCAGCTTAATTTGGC 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
451 CTCTCCACTCTGTCTCAGTATGATTAATTTTCTCTGAAAGAGCCCTTCAAG 500
167 gLysAsnProThrArgTrpSerSerThrAlaGlyPheLeuLysGlnAsn 184
501 GAAAAATCCACACGCTGCTCAACAACAGCTGCTCTCCTCAAGCACAACA 550
184 hAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTyrArg 200
551 CTGCTGGGAGTCTCTAAMACCTGGACAGAGAG..... 582
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
582 582
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234
582 582
234 IuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
583GATGAGATGGGACAGAAAGAGATTAACAGTCTGT 618
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
619 GAACCGTTGGACATGCTGACACGGGTTGGAGCATATATACCAACTTTTC 668
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
669 TCTGATGATATGATTAAGCTCTGTAAGTCCCAACGATGAGGGCTTC 718
284 euGlyIleHisValValProPheSerAlaArgGlyGlyValTrpThrLeuGly 300
719 TGGGAATCCATGATAGTCTTCAAGTCTGAGGGCGGAGAACCTTGCGG 768
301 LeuLeuValLysArgLeuGlnLysGlyGlyLysAlaGlnHisGluAsnLe 317
769 TTATTATGTAAGAAAGATGGAGAAAGGTGTAAGCTGAACATGAATAATGT 818
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
819 TTTTGTGAGATGATGATGCTGTCAGGATTAATGATGGCGACCTTCGAA 868
334 snArgArgPheGlnGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350

869 ATGAGATTTTGAACAAGCACACATATGTTTCCGCAAGCATGCGTACA 918
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrG 367
919 CCCATATTTGTTCCATGTGTTCTGTGACCAAAATTAAGACGATATGA 968
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
969 ACAACTATCCCAAAAGTAGAAGAACATTAATTTCAAGCCGTTTATGCC 1018
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1019 CTGACAGACAGATATATTGACAAACAGAGAGTGTGAACAGTGCAGGCTTCA 1068
401 ThrValGlnArgAlaProArgLeuAsnHisProGluGlnIleAspSe 417
1069 ACGGTCCAGAGACCCCGACTGAACCAACCCCTGAGCAGATTAACATC 1118
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1119 TCACCTAAGACTACCTCATAGGACACACCCCTGGGAAAAACCAATCCG 1168
434 IAProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1169 CTCACACCTCGGCACCTCAGAAATGTAATTAGTACACTGTAAACAGTGT 1218
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1219 TATTAACCAAAAAAATAGGCAAGAGCTTAATATCCAGTTAAAGAAAG 1268
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1269 TACAGAAAGTTTGGATTCAGCATCTTCACAGATGTAACATATAGTG 1318
484 IySerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaIle 500
1319 GCTCACCTCCAATCTATGTGAAMACATCTCCCGGGGGCGGCATAT 1368
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGlnValAsnGlyVa 517
1369 CAGGATGGCCGACTTAAGGCGAGACAGCTTAATAGAGTAATATGAGT 1418
517 IAspLeuValGlyLysSerGlnGluGlnValValSerLeuLeuArgSerT 534
1419 AGATTATAGTGGCAATCCCAAGAGGAAGTGTGTTGCTGTGTAGAACCA 1468
534 hTrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
1469 CCAAGATGGAGAGAACTGTAGCCCTTGTGCTTTCGCAAGAAAGAGCC 1518
551 PheHisProArgGlnLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
1519 TTCCACCCCAAGGGAACCTG..... 1536
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThrA 584
1537AAGCAGAAGATGAGGATTTGTTCTTACACTGATGGACCA 1579
584 rGluLeuPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
1580 GGAATTTCTGACATTTGAATGCCACTTAATGATTCAGAGATCTGCAGGC 1629
601 LeuGlyValSerValLysGlyLysAsnArgSerLysGluAsnHisAlaAsp 617
1630 CTTGGTGTAGTGTCAAAAGTAAACCGTCAAAAAGAACACCAACCAATTT 1679
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
1680 GGGAACTTTGTCAGTCCATTAATTAATGAGAGAGACGATCTAAAGATG 1729
634 IyArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 650

1730 GAAGCTTCGGTGAATGATCACTAGACACTAAATGAGAAATCCCTG 1779
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgSerMetSe 667
1780 TTGGCCAGCAACAACCAAGATGCCATGGAAACCTTAAGAAAGCTATGTC 1829
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
1830 TACTGAAGCGAAATAAACGAGCAATGATCCAGCTTAATGTTGCAAGGAA 1879
684 IeSerLysCysAsnGlnLeuLysSerProGlySerProGlyProGlu 700
1880 TAAGCAAGTCAATGAGCTCACTACCTGGAGCCCCCTGGACCTGAG 1929
701 LeuProIleGluThrAlaLeuAspArgGluArgArgIleSerHisSe 717
1930 CTGCCCATTTGAACAGCGTTGGATGATAGAGAAAGAAATTTCCCATTC 1979
717 rLeuLysSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
1980 CCTCTACAGTGGATGAGGGGCTTGATGTAATCCCCAGCAAAATGCTG 2029
734 IeLeuSerArgIleMetGlyLysSerGlyLysThrGlnLeuSerProThr 750
2030 CCTCAGTAGAGTAATG.....GGTAATACCAAGCTGCTCCCTACA 2070
751 ValAsnMetProGlnAspPThrValIleIleGlnAspArgLeuP 767
2071 GTGAATATGCCCAAGATGCACACTGATATATAGAAAGATGACAGGTGCC 2120
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAsp 784
2121 AGTGTCTCCACACATCTCTGACCAAGTCTCTCCAGCTCCCATGATG 2170
784 sValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
2171 ATGTGGGGTTTGGACGGCAGATGCTGTGTTGGGCCAAGGCTGCATTC 2220
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
2221 AGTATTTACACCGACGCTCTTGTAGTCCAGATGTTGATCCAGTCTTGGC 2270
817 aPheGlnArgGlyGlyPheGlyArgGlnSerMetSerGlyLysArgThrL 834
2271 TTTTTCACAGGAAGATTGGAGCTCAGACTATGTCAGAAAAACGCCCAA 2320
834 YsGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
2321 AGCAATTTTACATGCCATGCAATTTGATTTCTTAAACACGAAATCA 2370
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
2371 AAAAGCATGGATTAGT..... 2388
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2389AGTCTCCACAGACAGATGTGGGTCTTCCCTGGGCTC 2425
884 eutLysLysSerSerLeuGlnSerLeuGlnThrAlaValAlaGlyVal 900
2426 TGAAGAAGTCAAGCTCTTGGAGAAGTGTGCAGACCGCATGGCCGAGTG 2475
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleLeu 917
2476 ACTTTGAATGGGATATTTCTTTCCATGTCACAGCGCCGCGAATAACAG 2525
917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerYr 934
2526 AGCGAGGGGATGCAATAGAGCTTCAGAGCTCCATCGACAAAATCTTATG 2575
934 sPlysProAlaValAspAspAspAspGluLysMetGluThrLeuGluGlu 950
2576 ATAAACCCGGGTAGATGATGATGATGAAGCATGAGACCTTGGAAAGA 2625

951 AspThrGluLysSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
2626 GACACAGAAAGAAAGTTCAAGATCAGGAGAGAGCTGTATTCACACAGCAG 2675
967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGlu 984
2676 TGATCACCTTCCCTCTCTGTGAGAGACAAATGAATGAAGAAACCAAGA 2725
984 YsGlyAspLysThrAspArgLysAspLysThrGlyLysGluLysLys 1000
2726 AAGGTGTAAGACTGATAGAAAAAGCATAAACTGAAAAAGAAAGAAAG 2775
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
2776 AAGATAGAGATAGAGAGAAAGATAAATGAAGAACCAAGAAAGGAAATGCT 2825
1017 uLysGlyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspPL 1034
2826 GAAGGCTTGGAGACATGTTCC..... 2847
1034 YsIleGluLysThrGlyLysIleLysIleGlnGlnLysSerPheThrSerGlu 1050
2847 2847
1051 GluGluArgIleArgMetLysGlnGluGlnGluArgIleGlnAlaLysTh 1067
2848AGCATTCACAGCAAAAC 2864
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTrpAlaGluIleG 1084
2865 TCGAAGATTTAGAGGACGACAAAGCTGAGAGCGTACCTATGCTGTAATTC 2914
1084 IAspPheHisArgThrPheGlyCysAspAspGluLeuMetYrGly 1100
2915 AAGATTTTCATCGGACATTTGGCTGTGATGATGATTAATATAGGGGA 2964
1101 ValSerSerTyrgLysSerMetAlaLeuAsnAlaArgProGlnSerPr 1117
2965 GTTCTCTTATGAAGGTTCCATGCTCAACGCTGACCTGACAGGCC 3014
1117 oArgGluGlyHisMetMetAspAlaLeuTrpAlaGlnValLysLysProA 1134
3015 ACGGAAGGGCATATGATGATGCTTTGTATGCCCAAGTCAAGAACGCCG 3064
1134 rAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
3065 GGAATTTCCAAACCTCACCTGTAGACAGTAACAGATCAACTCCTAGCAAT 3114
1151 HisAspArgIleGlnArgLeuArgGlnGluPheGlnGlnAlaLysGlnAs 1167
3115 CATGATCGATACAGCGCTGTGAGCAAGAAATTTACAGCAAGCAAGCAAGA 3164
1167 pGluAspValGluAspArgArgArgTrpThrYrSerPheGluGlnProTrpP 1184
3165 TGAAGATGTAGAAAGATGTCGGCGACCTATATGTTTGAAGAACCCCTGGC 3214
1184 rAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
3215 CGAAGCGACGGCGCGGACGACGAGAGCGCGGCACTCGCTGCTGGAG 3264
1201 ValGlnMetGlnArgGlnArgGlnGluGluArgLysSerSerGlnGlnAl 1217
3265 GTGCAGATGCGACGGCGACGCGAGAGAGCGCAAGACCTCCACAGAGCG 3314
1217 aGlnArgGluTrpSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234
3315 CCAGCGCCAGTACAGCTCTGTGCTCGGCAAGAGAGAAAAATCCCACT 3364
1234 eValSerGlnAspSerTrpGluGlnAsnTrpSerProGlyGluGlyPhe 1250
3365 CGGTCTCCAGACTTGTGGAGCAGAACTACTCCCTGGGGAAGGCTTC 3414

710 CTATCCACTTCGTCAGTATTAACAACCTTTCCTCAGAGAGAGCCCTCCAG 759
167 glycAsnProThrArgTrpSerThrIraIaGlyPheLeuLysGlnAsnT 184
|||||
760 GAAAAACCCACCGCGTGGTCCAGCAGACAGCTGGCTTCTCAACAGACAGA 809
184 hIraIaGlySerProLysThrCysAspArgLysAspGluAsnTyrArg 200
||:|
810 CCACTGGAGTCCCAAAACCTCGCAGACAGAAAGATGAAACTACAGC 859
201 SerLeuProArgspThrSerAsnTrpSerAsnGlnPheGlnIraArgspAs 217
|||||
860 AGCCTCCACGGGATCCAGTACTGGTCCACACAGTCCAGAGAGACAA 909
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeug 234
|:|:|
910 CGCCCGCTCTCTGTAGTCCAGCCACCAATGTAAGACCGGTGGTGG 959
234 LuLysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250
|||||
960 AGAAGCAGAAACAGACAGAGAAAGGCACAGAAAGACACAGCCGAGTGT 1009
251 GluProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhse 267
|||||
1010 GAGCCGTTGGACATGCTGATCTGGAGTGGAGATATGATGCCAACTTTC 1059
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLysProL 284
|||||
1060 CCTCGATGATATGTAAAGCTCTACAAAGTCCCAACGATGGAGGGGCC 1109
284 euGlyIleHisValValProPheSerAlaArgGlyIleArgThrLeuGly 300
|||||
1110 TGGGAATCCATGTAGTGCTTTCAAGTCTCGAGCGGCAGAACACATGGGG 1159
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisIleAsnLe 317
|||||
1160 TTTTGTGACAGCGGTTGGAAAGGTGGAGAGCTGAACAGAAACCT 1209
317 upheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
|||||
1210 TTTCCAGCAGAAATGACTGCATGTGAGAGATTAAAGATGGAGATCTTGAA 1259
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
|||||
1260 ACAGAGATTGTGAGCAAGCACACCATATGTTCGCGCAACCATGCGTGGC 1309
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTyrG 367
||:|
1310 CGAGTCAATTGGTCCACGTGGTCCCTGCGACGACAAAGAGACGATATGA 1359
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
|||||
1360 ACAGCTCTCCCAAGCAGAGATGAACAATATTCCTCCGCGGTTCAGCC 1409
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerIaGlyLeuHis 400
||:|
1410 CCAGCAGCCACTGTGTGGCCACAGAGAGTGTGGCCAAATGCGCCCTCAG 1459
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
||:|
1460 GCGTTGCCAGAGACACCCAGACTGAGCCACACCTGAGCAGCGTGGAGTGC 1509
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProserA 434
||:|
1510 TCACCCCGGACTACCTCATGTGCTCATGCTCAACCAACCGGCCACAG 1559
434 laProAlaSerAlaProGluAsnValPheSerThrThrValSerSerGly 450
|||||
1560 CCCACGCTTGGCTCCACCAAGTGTCTAGTACACAGTGTGCGAGTGTG 1609
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
|||||
1610 TACAAACAGAAAGAGTAGGCAAGAGGCTCAACATTCAGCTTAAAAAAG 1659

467 yThrGluGlyLeuGlyPheSerIleThrSerArgspValThrIleGlyG 484
|||||
1660 TACAGAAAGACTGGGATTTAGCATCACTTCCGAGAGACTATATAGT 1709
484 ySerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaIle 500
|||||
1710 GCTCAGTCCCATTTATGTCAAGAACATTTCTTCTCCAGGGGGTGCATTT 1759
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
|||||
1760 CAGAGTGGCAGGCTCAAGGCGGAGACCGGCTATATAGAGTCAATGGAGT 1809
517 laAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
||:|
1810 AGATTATGCGGCGCAAGTCCAGAGAGAAAGTGTTCCTGTGGAGAAACA 1859
534 hIryLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
|||||
1860 CCAAGATGAGGGGACCGTGAAGCTTGTGGTCTTTCGTCAAGAAAGAGCT 1909
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
|||||
1910 TTCCACCCAAAGGAAATGAATGCCGAACCCAGCCAGATGCCAGACTCCAAA 1959
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
|||||
1960 AGAAACAAAGCCGAACAGCAGACGATTTGTTCACACCTGAGCGTACCA 2009
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
|||||
2010 GGGGTTTCTGACTTTCGAAGTTCCTCACTCAATGATCAAGGTCGACAGG 2059
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
|||||
2060 CTTCGTCTACGCTCAAGGGGACCGCTCCAAAGAAACCATGCGAGATT 2109
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
|||||
2110 GGGGATCTTGTTAAATGTCATCATTAACGGGGGAGCTGCATCTAAAGATG 2159
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
|||||
2160 GAGAGCTGAGAGGTAAATGACCAGCTGATAGCTGGAACGGAGAGTCTCTG 2209
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
|||||
2210 CTGGGCAAAAGCCAAACGAAGACCATGAGACTGTGAGAGAGTCAATGTC 2259
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
|||||
2260 CACGAGAGGCAACAAGCGTGGCATGATCCAGCTCATGCTGGCGAGGGGGA 2309
684 leSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
||:|
2310 TCACCAATATGAACGAGTTCGGGTCTCTGTGGACCCCGCGCACCCGAG 2359
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
|||||
2360 CTGCCCTAGAACACAGAGATTGGATGACAGAGAACGCAAGATCTCACACTC 2409
717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAlaA 734
|||||
2410 CCTTACAGTGGGATTCGAGGGGCTGAGTGAAGTCTCCACACGAGATGCCG 2459
734 laLeuSerArgIleMetGlyLysSerGlyLysTyrGlnLeuSerProThr 750
|||||
2460 CACTCACACAGGATTAATGGTGAGTCAAGTAAATCCAGCTCTCCCAAC 2509
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
|||||
2510 GTGAACATGCCACATGATGACACTGTCTATGTGAAGTGAACAGAGTGGCC 2559

```

767 oValLeuProProlHisLeuSerAspGlnSerSerSerHisAspA 784
|||||
2560 TGTGCTCCCTCTCACTCTGACCAAGCTCTCCAGCTCCATGATG 2609
784 spValGlyPheValThrAlaAspAlaGlyThrTriPAlaLysAlaIle 800
|||||
2610 ACCTGGGATTCATTAATGACAGAGCGACGCTGGCCAGGCTTACCATC 2659
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
|||||
2660 AGTGACTCAGCCGACTGCTCATTTAGTCCAGATGTTGATCCGTTCTCGC 2709
817 abPheGlnArgGlyGlyPheGlyArgGlnSerMetSerGlyLysArgThrL 834
|||||
2710 TTTTCAACAGAGAGGATTTGGACGCCAGATATGTAGAAAAACGACAA 2759
834 yslGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
|||||
2760 AGCAATTTTCAATGCCAGTCAATTTGCTTAAACAGCAAAATCA 2809
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
|||||
2810 AAAAGCATGGATTAGTATAGCTGACGAGACTMAACTCATATCACTGGA 2859
867 paAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
|||||
2860 TGACCGAGAGGACAGGCTCCCCCAATGAGATGTGGGACCTCTTGGGTC 2909
884 euLysLysSerSerSerLeuGlnSerLeuGlnThrAlaValAlaGluVal 900
|||||
2910 TGAAGAAATCCAGCTCTTWTAGAAAGTCTGACAGCGCTTGTCTGAGTGC 2959
901 ThrLeuAsnGlnLysPheProPheHisArgProArgProArgIleIleArg 917
|||||
2960 ACCCTGATGGGAAACATCTTCTTCCACCGCCACGGCCAGCAATCATCCG 3009
917 gGlyArgGlyCysAsnGlnSerPheArgAlaAlaIleAspLysSerTyrA 934
|||||
3010 AGGAGAGGGCTGCACAGAGCTTCAGAGCGGCATTTGACAAAGTCTACG 3059
934 spLysProAlaValAspAspAspAspGlnGlyMetGluThrLeuGluGlu 950
|||||
3060 ATAAAGCCCATGTGATGACGACAGCAAGCATGGAGACCTTGGAGAA 3109
951 AspThrGlnGlnSerSerArgSerGlyArgGlnSerValSerThrAlaSer 967
|||||
3110 GACACAGAAAGAAAGTCAAGGTCAAGGAGGAGATCCCTGTCCAGTCCAG 3159
967 rAspGlnProSerHisSerLeuGlnArgGlnMetAsnGlyAsnGlnLut 984
|||||
3160 TGATCAGGCTTCTTAATCTCTGAGAGCAAAATGAAATGAGAACCCAGAGA 3209
984 yslGlyAspLysThrAspArgGlyLysAspLysThrGlyLysGluLysLys 1000
|||||
3210 AGAGGGAGCAAGAGAGAGAAAAAGCAAAAGCCGGAAAGATTAAGAG 3259
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
|||||
3260 AAAAGCCAGAGAAAGAGAAAGATTAATGAAAGCCAAAGAGGGGATGCT 3309
1017 uLysGlyLeuGlnLysPheMetPheArgPheGlyLysHisArgLysAspL 1034
|||||
3310 GAAAGGCTTGGGGGACATGTTCAAGTTTGGCAAAATCGAAAAAGATGCA 3359
1034 yslIleGlnLysThrGlyLysLysLysLysLysLysLysLysLysLysL 1050
|||||
3360 AGATGAGAAAAATGGCTGAAATTAATTAATCCAGACTCTTTTCACTCAGAA 3409
1051 GluGluArgGlyLeuArgMetLysGlnGlnGlnLutArgLysGlnAlaLysTh 1067
|||||
3410 GAGGACAGAGGCTCGGATGAGAGAGAACAGAGGATTCAAAGCCAAAC 3459
1067 rArgGluPheArgGluArgGlnAlaArgGluArgAspTyrIleArgLutIleG 1084
|||||
3460 TCGAAGATTTTAGGAGCGGCAAGCTCGAGAGCCGACTAATGTAATGCAATCC 3509
1084 lnaAspPheHisArgThrPheGlyCysAspAspGluLutMetTyrGlyGly 1100
|||||
3510 AAGATTTTCCATCGAACAATTTGGCTGTGTATGACGAGTTGCTGATGGGGG 3559
1101 ValSerSerTyrGlnGlySerMetAlaLeuAsnAlaArgProGlnSerPr 1117
|||||
3560 ATGTATCTCTTATGATGGCTGGCTGCTCTCAATGCCAGACCCAGAGCCCC 3609
1117 oArgGlnGlyHisMetLeuAspAlaLeuTyrAlaGlnValLysLysProA 1134
|||||
3610 AAGAGAAAGTCACTCATGATGACACTTTGTATGCACAAAGTAAAGAAACCTC 3659
1134 rGAsnSerLysProSerProValAspSerAsnArgSerThrProSerAsn 1150
|||||
3660 GGAGCTCCAAACCTGGC.....GACACCAATTCATCCACCCTTAGCACAC 3703
1151 HisAspArgLysGlnArgLeuArgGlnGlnPheGlnGlnAlaLysGlnAs 1167
|||||
3704 CATGACCGGATATCAACGCTTACGGCAAGAAATTCACAGAACCCAAACAGA 3753
1167 pGluAspValGluAspArgArgArgThrTyrSerPheGlnGlnProTrpP 1184
|||||
3754 TGAGACGCTGGAAGACGCGCGCTACCTACAGCTTGAAGCATCTTGTCT 3803
1184 roAsnAlaArgProAlaThrGlnSerGlyArgHisSerValSerValGlu 1200
|||||
3804 CCAGCTCAAGGCGCACATCGCAGATGTGTGCGCACTCGGTGCAATGAGAG 3853
1201 ValGlnMetGlnArgGlnArgGlnGlnGlnGluGlnLysSerGlnGlnAl 1217
|||||
3854 GTTCAAGTGCACGCGGACGCGCGCAGAGAGGAGCGAGAGCTTCCACAGGC 3903
1217 aGlnArgGlnLysSerSerLeuProArgGlnSerArgLysAsnAlaSerS 1234
|||||
3904 CCAAGCGCAGTACAGTCCCTCTCCAAAGACAAAGACAGAAATGCCAGCT 3953
1234 erValSerGlnAspSerTrpGlnGlnAsnTyrSerProGlyGluGlyPhe 1250
|||||
3954 CTGTATCACAGATTCCTGTGGAACAGAACTACGCCCTGTGAAAGCTTTC 4003
1251 GlnSerAlaLysGlnAsnProArgTyrSerSerTyrGlnGlySerArgAs 1267
|||||
4004 CAGAGTGCACAGAGAAACCCAGGATTTCCAGTTACAGAGGCTCCAGGAA 4053
1267 nGlyTyrLeuGlyGlyHisGlyPheAsnAlaArgValMetLeuGluThrG 1284
|||||
4054 CGGCTATCTAAGTGGCCATGGCTTTAATGCCAGGGTCAATGCTGAGACCC 4103
1284 lnsLutLeuLeuArgGlnGlnArgAlaArgLysGlnGlnGlnMetLysLys 1300
|||||
4104 AGGAGCTTTCGCCAGACAGAGACAGCGCAAAACAGCAAGATCGAAGAAG 4153
1301 GlnProProSerGlnGlyProSerAsnTyrAspSerTyrLysValGln 1317
|||||
4154 CAGCCACCAAGGAGTGA..... 4171
1317 naSProSerTyrAlaProProLysGlyProPheArgGlnAspValProp 1334
|||||
4172 .....GTCAAGAGGGCCCTTCCGGCAAGATGATGCCCC 4202
1334 roSerProSerGlnValAlaArgLeuAsnArgLeuGlnThrProGluLys 1350
|||||
4203 CGTCTCATCTCAGGTGGCCAGGCTGAACAGACTGCAGACACCCGAGAAA 4252
1351 GLyArgProPheTyrSer 1356
|||||
4253 GGAGCGGCTTCTTACTCTC 4270
seq_name: gb_ro:AY026057

```

seq_documentation_block:
LOCUS AY026057 4167 bp mRNA linear ROD 20-FEB-2001
DEFINITION Mus musculus PAR-3 180 kDa isoform (Par3) mRNA, complete cds.
ACCESSION AY026057
VERSION AY026057.1 GI:12965348
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 4167)
TITLE Lin.D., Gish.G.D., Songyang,Z. and Pawson,T.
The carboxyl terminus of B class ephrins constitutes a pDZ domain
binding motif
JOURNAL J. Biol. Chem. 274 (6), 3726-3733 (1999)
MEDLINE 99121117
PUBMED 9920925
REFERENCE 2 (bases 1 to 4167)
AUTHORS Lin.D., Edwards,A.S., Fawcett,J.P., Mbanalu,G., Scott,J.D. and
Pawson,T.
TITLE A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and apKC
signalling and cell polarity
JOURNAL Nat. Cell Biol. 2 (6), 540-547 (2000)
MEDLINE 20394297
PUBMED 10934475
REFERENCE 3 (bases 1 to 4167)
AUTHORS Lin,D.C. Jr. and Pawson,T.J.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2001) Programme in Molecular Biology, Samuel
Lunenfeld Research Institute, 600 University Avenue, Toronto,
Ontario M5G 1X5, Canada
FEATURES
source Location/Qualifiers
1..4167
/organism="Mus musculus"
/strain="NIH Swiss"
/db_xref="taxon:10090"
1..4167
/gene="Par3"
110..4111
/gene="Par3"
/note="mpar-3: scaffolding protein"
/codon_start=1
/product="PAR-3 180 kDa isoform"
/protein_id="AAK07669.1"
/db_xref="GI:12965348"
/translation="MKVTVCGRTRVVVPCGGRMKVPSLIQOAVTRYRKAVAKDPNY
WIOYRLRHGGGILDDLDICDVADDDKRLVAVVEDDDPHGGGCTASPTGQSP
IFGSEICNNYSARQPYOATSEIEVTPSYLRAMPILYRRSRDPLITGLTSYNNF
SSEPSRKNPTRMSTTAGFLKONTAGSPKTCRKDEYRSLPRPSSMSNOFQDNA
RSSLSASHPNVDRMLEKODEEGTEEDSRYEPYHADTLENPNFSLDMVKLYO
VPNDGPGIIVHPFPGSARGGRTILGLVRLKRGKAEQENLFHENDCLVRIINDGLRN
RRFEQAHMFRQAMRAVIMFHVYPAKKEQYEOLEKKNYSFGSPSYSTNGVNR
SVANNAPOALPRAPRLSQPQLDAPRLPHSAHASTKPPAPALAPSYSTNGSV
YNTKKVGRILQIKKKEGIGFSTSDYITGSAPIYVNIILRGAILQDGRKAG
DRLTEVNGVLAGKSQEEVSLRSTKMEGVSLVTFQEEAFHRENNAEESQMQPT
KETAEDEVDVLPDGTREFLTEVPLNDSGAGVGAVKGRSKENADLIEFKSI
INGGAASKDGLRLVNDQLIAVNGESILGKANDAEETLRSTSTGNRMAQLIVAR
RISRCNELRSGSPAPAPLPIETELDERRRSHSLYIEGLSDSPRNALSTRIMG
KCOLSPYVNMHDDVTMIEDRLPYLPHLSDOSSSSHDVGYITMFGATMKATIS
DSADCSLSPDYDYLVAROREGFRGKMSKEKTKOSDSQDLFYVTKRSKMDLYADE
TKLNTVDQRRGSPRDVPSGLKSSLESLOTVAVETLNGNIPHRPRLIRIG
RGCNESFPAALIDKSTDKPMVDDEGMETLEDEESSRGSSEYSTSDPSYLER
OMNGDPEKRDTERKDKKADKKDKREKDKLANKGMKLGIDMFRGFKHRSDK
MEKMGRIKIQDSFTSEEDRVNMEKEQERIQATREFRERARERDYAEIODEPHFGC
DDELLYGMSSEYEGCLALNARPOSPREGLMDTLYAYKKPPSSKPGDSNRETNHD
RIORROFOQAKODEVEDRRTYSFEOSWSSSPASQSGHSYSVEYQVROQOE
RESTRQARQYTSIPROSKNASSISQDSWENOTAPGGGFGSAKNPRTSSQSGRNG
YLGGEFNARVWLETQELLKQDQRRKEQOLAKOPPADVGRGFQDVPSPPSQVARN
RLQPEKRRPYS"

BASE COUNT 1145 a 1079 c 1174 g 769 t
ORIGIN

alignment_scores:
Quality: 6310.50 Length: 1356
Ratio: 4.858 Gaps: 4
Percent Similarity: 95.796 Percent Identity: 89.676
alignment_block:
US-09-757-781-2 x AY026057 ..
Align seg 1/1 to: AY026057 from: 1 to: 4167
1 MetysValThrValCysPheGlyArgThrArgValValAlaProCysGI 17
110 ATGAAGTGACCGTGTCTTGGAGGACCCGGGTGGTGGTGGTGGTGG 159
17 YASPGIYHISmetysValPheSerLeuILEGlnAlaValIThrArgT 34
160 AGATGCCCGCATGAAGTTTTCACCTCATTCACAGCGGTGACCCGCT 209
34 YrArgValAlaIleAlaLysAspProAsnTYRTPILeglnValHisArg 50
210 ACCGGAAGCGCGTGCCCAAGGATCCAAACrACTGATACAGTGCATCGC 259
51 LeuGlnHisGlyAspGlyGlyIleLeuAspLeuAspAspIleLeuGly 67
260 TTGGAGCATGGAGATGGAGGATTTCTAGACCTGGATGACATCTTGTA 309
67 PValAlaAspAspLysAspArgLeuValAlaValPheAspGlnAsp 84
310 CGTTGCTGATGACAAAGACAGACTGTGTAAGTTTTCATATACAGATC 359
84 roHISHisGlyLysAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
360 CCCACCATGGAGAGATGTGTACAGCGGCACCTTCACGGGAACCCAGT 409
101 ProGluIlePheGlySerGluLeuGlyThrTrpAsnValSerAlaPheGI 117
410 CCAAGATATTGGGACGTGAGTGGGACCAACAATGTTCTTGCTTTTCA 459
117 nProTyGlnAlaThrSerGluIleGlnValThrProSerValLeuArg 134
460 GCCTTATCAACCCCAAGTGAATTTGAGGTACCCCTTCACTTTCGGG 509
134 LaaSerMetProLeuHisValArgArgSerSerAspProAlaLeuILEg 150
510 CAATATGCGCTTTCATGTCCGCGGAGACAGACCCAGCTTAACTGGC 559
151 LeuSerThrSerValSerAspSerAsnPheSerGluGluProSerAr 167
560 CTTCACCTTCTGTCACTGTATMACACTTTTCTCAGAGGAGCCCTCCAG 609
167 GlyAsnProThrArgTrpSerThrAlaGlyPheLeuLysGlnAsnT 184
610 GAAAAACCCACCGCGTGTCCACAGACAGCTTCTTCTCAAGCAGAA 659
184 hralaGlySerProLysThrCysAspArgLysLysAspGluAsnTYrArg 200
660 CCGGTGAAGTCCCAAAACCTGCGACAGAGAAGAAAGATGAACATACGA 709
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnInPheGlnArgAs 217
710 AGCTTCCACGGGATCCAGTGTGTCACCAACAGTTCACGCGACACA 759
217 nAlaArgSerSerLeuSerAlaSerHisPrometValGlyLysTrpLeuG 234
760 CGCCGCTCCTCCTTGAACCGCACACCAATGATGACGCGGTGCTGG 809
234 LuIysGlnGluGlnAspGluAspGlyThrGlnGluAspAsnSerArgVal 250
810 AGAAGCAAGAAACAGATGAGGAAGCACAGAAAGAACAGACAGCGAGTG 859
251 GluProValGlnHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
860 GAGCGGTGGACATGCTGATACCGGATGGAGAAACATGCCCAACTTTC 909

```

267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyProL 284
    |||||||
910 CCTCGATGATATGTAAGCTCTGACAACTCCCAACGATGAGGGCCCC 959
284 euGlyIleHisValIleProPheSerAlaArgGlyIleArgThrLeuGly 300
    |||||||
960 TGGGAATCATCTAGTGGCTTTCAGTCTCGAGCGGCGCAACAATTGGGG 1009
301 LeuLeuValLysArgLeuGlyGlyGlyLysAlaGluHisGluAsnLe 317
    |||||||
1010 TTGTAGTGAAGCGGTTGAGAAAGCGGTAAGCGCTGACCAAGAAACCT 1059
317 uPheArgGluAsnAspCysIleValIleArgIleAsnAspGlyAspLeuArg 334
    |||||||
1060 TTTCATGAGAAATGACTGCACTGTGTGAGATTACATGATGAGATCTGCA 1109
334 snArgArgPheGluGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
    |||||||
1110 ACAAGAAATTTGAGCAAGCAACAATATGTTCCGCAAGCTATGCGGCG 1159
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnIleArg 367
    |||||||
1160 CGTGTATTTGGTTCATGTGTCTCCCTGCAACAACAAGACCAATATGCA 1209
367 uGlnLeuSerGlnSerGluLysAsnAsnTyrTyrSerSerArgPheSerP 384
    |||||||
1210 ACAACTGTCACAGCGGAGAAACAACACTACTCCCGAGCGCGCTTTCAGC 1259
384 roAspSerGlnTyrIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
    |||||||
1260 CTGACAGCCACTGTGTGGCCACAGAGCTGTGGCCAAACAATCCCCCTCA 1309
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
    |||||||
1310 GCATTTGCCAGACGCCACCAAGCTGAGTCAGCCACCGAGCAGCGGAGTGC 1359
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProSerArg 434
    |||||||
1360 TCACCCCGACTACCTCATAGTGTCTACGCTCAACCAACCAACCCCGCAG 1409
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
    |||||||
1410 CCCCGCGCTTGGCTCCACCCAGTGTGTAGTACCAACGTAAGCAATGTC 1459
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysSerG 467
    |||||||
1460 TCACACACGAAAGAAAGAGCAAGAGGCTCAACATCAGCTTAAGAAAG 1509
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
    |||||||
1510 TACAGAAAGAGCTGGATTTCAGCATTCCTCCGGGACGTCACCAATAGGTG 1559
484 LysSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaIle 500
    |||||||
1560 GCTCAGCTCCCATTTATGTCGAAGAATATCTTCGAGGGGCTGCCATTT 1609
501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517
    |||||||
1610 CAGGATGGCAGACTCAAGCAGAGACCGGCTAATAGAGGTCAATGGAGT 1659
517 laSpleuValIleLysSerGlnGluValValSerLeuLeuArgSer 534
    |||||||
1660 AGATTTAAGCAGCAATCCCAAGAGGAAGTGTTCCTGTTGGAAGCA 1709
534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
    |||||||
1710 CCAAGATGGAGGAGCTGTGAGCTTCTGCTTCTCAGGAAAGCGCT 1759
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
    |||||||
1760 TTCCACCCCAAGGAATGATGCTGAACCAAGCCAGATSCAGATCTCCAAA 1809
567 sGluThrLysAlaGluAspGluAspIleValIleThrProAspGlyThr 584
    |||||||
1810 AGAAACGAAAGCTGAAGATGAGAGCTGTCTCACACCCGATGATCA 1859
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
    |||||||
1860 GGGAGTTTCGACTTTTGAATCTTCACATGATGATCAGAGATCTGCAGGG 1909
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
    |||||||
1910 CTGTGTGCTGACTGTCAAGGGGAAACCTTCCAAAGAAACACAGCGAGATT 1959
617 uGlyIlePheValLysSerIleIleAsnGlyAlaAlaSerLysAspG 634
    |||||||
1960 GGGGATCTCGTTAAATTCATTATCATGATGAGACTCATCTAAAGATG 2009
634 LysArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysSerLeu 650
    |||||||
2010 GAAGCGTGAAGGCTAAATGACAGCTGATACCTGTCAATGAGAACTCTCTA 2059
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
    |||||||
2060 CTGGCAAAAGCCACACAGAAAGCCATGAGACTCTACGAGGCTCATGTGC 2109
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArg 684
    |||||||
2110 CACCGAGGCAACAAGCGTGCATGATTCACGCTCATTTGTGGCAGGCGGA 2159
684 leSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
    |||||||
2160 TCAGCAGATGTACAGAGCTCGGCTCTCTGAGCCCTCTCCACTAG 2209
701 LeuProIleGluThrAlaLeuAspAspArgGluArgGlyLeuSerHisSe 717
    |||||||
2210 CTGCCCATCGAGACAGAACTGATGACCAAGACGACGATCTCACACTC 2259
717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAla 734
    |||||||
2260 CCTCTACAGTGGGATGAGAGGCTGTGATGAGTCGCCACCAAGAAACGAG 2309
734 laLeuSerArgIleMetGlyGluSerGlyLysTyrGlnLeuSerProThr 750
    |||||||
2310 CACTGAGCAGAGTAATG.....GCTAAATGCCAGCTGTCCTCCACG 2350
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuP 767
    |||||||
2351 GTGAACATGCTCAGATGACATGATCATGATGAAAGATGACAGGCTGCC 2400
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerSerHisAspA 784
    |||||||
2401 TGTGCTCCCTCTCACCCTCTGTGACAGTCTCTCCACTCCCATGATG 2450
784 spValGlyPheValThrAlaAspAlaGlyThrTrpAlaLysAlaIle 800
    |||||||
2451 ACGTTGGATTCATATGACAGACAGCAAGCAAGTGGCCAAACCTACATC 2500
801 SerAspSerLysAspCysSerLeuSerProAspValAspProValIleAl 817
    |||||||
2501 AGTGATCAGCCGACTGCTCATTTGATGTCAGATGTGATCGGTTCTGCG 2550
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
    |||||||
2551 TTTTCAACGGGAAGATTGGACGCCACAGATATCTCAGAAAAAGCACA 2600
834 ySglnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
    |||||||
2601 AGCAATTTTCAGATCCAGTCAATTTGATTTCGTTAAACACCAAAATATA 2650
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
    |||||||
2651 AAAAGCATGATTTA...GTAGTGAAGAGACTAAACTCAATACAGTGA 2697
867 pAspGlnLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884

```


TITLE

Direct Submission

JOURNAL

Submitted (18-Oct-1999) Laboratory of Molecular and Cellular
Oncology, Shanghai Institutes for Biological Sciences, Chinese
Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, P.R. China

FEATURES

Location/Qualifiers

1..3470

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="10"

/map="10p11.2"

288..3383

/note="hasRpsa"

/product="atypical PKC isoform-specific interacting
protein short variant"

/protein_id="AAK27892.1"

/db_xref="GI:13491612"

/translation="MKYVCFGRVYVPCGGDHMKVSLIQAAVTRYKAIADPNV
IYSEELGNNVSAFOYATSEIYVPSYLRANMPLHVRSSDPALIGLSVSDSNP
RSLSASHPMKLEKRODDEGTEDNSRYEPYADGTGLHPIPSIDMKVLE
VPNDGGLGIVVPEFARGLTGLVLRLEKGAHEENYSSRPSDSOYDNR
RRPEOAHMFRAMPPIIMFHVPAKNEQYEOLOSSEKNNYSRPSDSOYDNR
YNRKIKGRNLNOLKMGTEGLFSTSDVTIGSAPRYVKNLIPGAAIODERLK
DRLIEVNGDLVKSQEEVSLRSTKMEGVSLIVRQEDAFHRELNAEPSONOIP
KETRAEDVDVLPDGRFELFEVPLNDSSAGLGVSVGNSKEVHDLGFYVSI
INGAASKDGRRLRVNDOLAVNGESILGKTMDMELTVRSNKEVHDLGFYVSI
KISKMEKSPSPGPELPJETALDDREKISLSISGLEDEGNKGMOLIVAR
KYOLSTVPMQDDVTIIEDRLVPLPHLSDSSHDVDFVATLADMAKAIIS
DSADSLSPVDVPLAFORBPGRQSSSEKRTYQFASDGLVTKRKSKNDLAD
ETRLNVDVDOKAGSPSPGGLSKSSLSLEIQTAVLEVLINDIPIPHRPRILIR
GRCNESFRALIDKSYDKPAVDDEMELEEDTESSRSRPSVSTASDPSISLE
ROMNGNOGKGTBRKDKTKERKKDKMAKKMGMLKGLDMFSLAKLPEK
R"

CDS

BASE COUNT	1003 a	811 c	934 g	721 t	1 others
ORIGIN					

alignment_scores:

Quality:	5268.50	Length:	1034
Percent Similarity:	99.323	Gaps:	1

alignment_block:

US-09-757-781-2 x AF196186

Align seg 1/1 to: AF196186 from: 1 to: 3470

```

1 MetLysValThrValCysPheGlyArgThrArgValValProCysG1 17
288 ATGAAAGTACCGTGTGTCGAGACGCCGGGTGTCGCGCGG 337
17 yAspGlyHisMetLysValPheSerLeuLeuGlnGlnAlaValThrArgT 34
338 GGAGGCGGCACATGAAGTTTCACCTCATCCACAGCGGTGACCCGCT 387
34 yAlaGlyValAlaIleAlaLysAspProAsnTyTrpIleGlnValHisArg 50
388 ACCGGAAGCCATCCGCAAGATCCAAACATCTACGTATACAGGTGACATCC 437
51 LeuGlnHisGlyAspGlyGlyIleLeuAspLeuAspPrlLeuCysAs 67
438 TTGGAACTGAGAGTGAAGAACTCTGACCTTGATGACATCTTTGTGA 487
67 yValAlaAspAspLysAspArgLeuValAlaValPheAspGlnGlnAsp 84
488 TGTAGCAACGATTAAGACACACTGCTAGCAGTGTGATGACAGAGATC 537
84 roHSHISGLyGlyAspGlyThrSerAlaSerSerThrGlyThrGlnSer 100
538 CACATCAACGAGGTGATGACACAGTTCACCGGTATCCACAGAGC 587

```

```

101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
588 CCAGACATATTGCTAGTACGCTTGGCCACCAACAAATGCTCACCCTTTA 637
117 nProTyGlnAlaThrSerGluIleGlyValAlaThrProSerValLeuArg 134
638 GCCTTACCAAGCAACCAAGTGAATTTGAGTGCACACCTTACGCTTCAG 687
134 IaAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
688 CAATATGCTCTTCATGCTTCGACCGCAGTGAAGTACCGCTCTAATTGCC 737
151 LeuSerThrSerValSerAspSerAsnPheSerSerSerGluGluProSer 167
738 CTCTCCACTCTCTCAGTATAGTATATTTTCTCTGAAAGCTTCAAG 787
167 GlyAsnProThrArgTyPserThrAlaGlyPheLeuIleGlnAsnT 184
788 GAAAAATCCACACACCTGTCMAACACACCTGCTTCTCAACACAGACA 837
184 hAlaGlySerProLysThrCysAspArgGlyLysAspGluAsnTyArg 200
838 CTGCTGGGAGTCTTAAACCTGCGACAGAAAGAAAGATGMAAACTACGA 887
201 SerLeuProArgAspThrSerAsnTyPserAsnGlnPheGlnAlaArg 217
888 AGCTTCCGCGGAGTACTAGTAACTGCTTAAACATTTTCAGAGACAA 937
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTyPleug 234
938 TGCTGCTGCTCTCTGAGTGCACAGTCCAAATGCTGGGCAAGTGCTGG 987
234 LuLysGlnGluAsnAspGlyLysGlyThrGluGluAspAspSerArgVal 250
988 AGAAACAGACAGACATGAGATGAGTGGACAGAAAGATTAACATCTGTT 1037
251 GluProValGlyHisAlaAspThrGlyLeuGlnHisIleProAsnPheSe 267
1038 GAACCTGTGGACATGTCGACACGGGTTTGACATATACCAACTTTTC 1087
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyGlyProL 284
1088 TCTGGATGATATGCTAAAGCTCTGACAAAGTCCCAACAGATGAGGCTTC 1137
284 euGlyIleHisValAlaProPheSerAlaArgGlyGlyArgThrLeuGly 300
1138 TGGATCCATGTTAGTGTCTTTCAGTGTGAGGCGGCAAACTTGGGG 1187
301 LeuLeuValLysArgLeuGluLysGlyGlyLysAlaGlnHisGluAsnLe 317
1188 TTTATGATAAAGATGATGGAGAAAGTGTGTAAGTGAATGAATCTT 1237
317 uPheArgGluAsnAspCysIleValAlaGlyLeuAsnAspGlyAspLeuArg 334
1238 TTTTCTGAGAAATGATTCATGTCATGATTAATGATGAGGACCTTCGAA 1287
334 snArgArgPheGluGlnAlaGlnHisMetPheArgGlnAlaMetArgThr 350
1288 ATAGAAATTTGAACAAAGCAACATATGTTTCCCAACCAAGTGTGACA 1337
351 ProIleIleTyPheHisValAlaProAlaAlaAsnLysGluGlnTyArg 367
1338 CCGATCATTTTGTGCTCATGTGTTCTTGCAGCAAAATTAAGAGCAGTATGA 1387
367 uGlnLeuSerGlnSerGluLysAsnAsnTyTrpTySerSerArgPheSer 384
1388 ACAACATATCCCAAGTGAAGAAACAATTAATTAATCAACGCGTTTATGCC 1437
384 roAspSerGlnTyTrpIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1438 CTGACAGCAAGTATATGACAAACAGAGTGTGAACAGTGCAGGCTTCAC 1487

```

```

401 ThValGlnArgAlaProArgLeuAsnHisProGluGlnIleAspSe 417
|||||
1488 ACGGTGAGAGAGACCCCGACTGAAACACCCGCTGAGCAGATAGACTC 1537
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
|||||
1538 TCACCTCAGACTACCTCATAGCCGACACCCCTGGGAAAAACCAACATCCG 1567
434 lArProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
|||||
1588 CTCACGCTGGGCACTCCAGATCTATTAGTACGACTGTAAAGAGTGGT 1637
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1638 TATATACACCAAAAAATAGGCAAGAGGCTTAATATCCAGCTTAAGAAGG 1687
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
|||||
1688 TACAGAAAGTTTGGGATTCAGCATCACTCCAGAGATGTAAACAATAGCTG 1737
484 lYSerAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
|||||
1738 GCTCAGCTCCAAATCTATGTAAAAACATTCCTCCCGGGGGGGCGCATTT 1787
501 GluAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVa 517
1788 CAGAGTGGCCGACTTAAGCGCAGGACAGACTTATAGAGGTAAATGGAGT 1837
517 lAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
|||||
1838 AGATTATGTGGCAATCCCAAGAGAAAGTTGTTCCCTGTGGAAGACA 1887
534 hTrLysMetGluGlyThrValSerLeuLeuValPheArgGlnGluAspAla 550
1888 CCAAGATGGAAGAGACTGTAGGCTTGTGCTTTCGCCAAGGAAGAGCC 1937
551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProly 567
1938 TTCACCCCAAGGGAACTGAATGCGAGACCAGCCAGATGCAGATTCCTCAA 1987
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
|||||
1988 AGAAAGCAAAAGCAGAAAGATGAGATATGTTCTTACACCTGATGGCACCA 2037
584 rGluPheLeuThrPheGluValAlProLeuAsnAspSerGlySerAlaGly 600
2038 GGGAAATTTCGACATTTGAAGTCCCACTTAATGATTCAGGATTCGACGCG 2087
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
2088 CTGGGTGTCAAGTGTCAAAAGGTAAACCGGTCAAAAGAGAACCAACGCAGATTT 2137
617 uGlyIlePheValLysSerIleIleAsnGlyValAlaAlaSerLysAspG 634
|||||
2138 GGGAACTTTGTCAAGTCCATTTAATTAAGAGAGGACGACATCTAAAGATG 2187
634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyLysLeu 650
2188 GAAGGCTTCGGGTGAATGATCACTGATAGCAGTAAATGGAGATCCCTG 2237
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
2238 TTGGGCAAGCAAAACCAAGATGCCATGGAAAACCTTAAGAGGTCTATGTC 2287
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
2288 TACTGAAGGCAATTAACGAGGATGATCCAGCTTATTGTTCAGAGAGAA 2337
684 lSerLysCysAsnGluLeuLysSerProGlySerProGlyLysProGlu 700
2338 TAAGCAAGTGAATGAGCTGAGATGCACCTGGGAGCCGCCCTGGAGCTGAG 2387
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
|||||
2388 CTGCCCAATTGAAMACGGCTTGATGATAGAGAACGAAGAAATTTCCCATTC 2437
717 rLeuTyrSerGlyIleGluGluLeuAspGluSerProSerArgAsnAla 734
2438 CCTCTACAGTGGGATTTGAGGGCTTGATGAATCCGCCAGCAAAATGCTG 2487
734 lAlaLeuSerArgIleMetGlyGluSerGlyLysTyrGlnLeuSerProThr 750
2488 CCCACAGTAAAGATTAATG.....GGTAAATACAGCTTCCCTCCATCA 2528
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
2529 GTCAATATAGCCCCCAAGATGACACTGTATATAGAAGATGACAGTTGCC 2578
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
2579 AGTCTTCTCTCCACATCTCTCTGACCACTCCTCTTCCAGCTCCCATATG 2628
784 sPValGlyPheValThrAlaAspAlaGlyThrTyrPalAlaLysAlaAlaIle 800
2629 ATGTGGGGTTTGTACGCGCAGATGCTGTGTTGGCCAAAGCTGCATTC 2678
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValLeuAl 817
2679 AGTGATTACGCGCAGCTGCTTGTAGTCCAGATGTGATCCAGTTCCTTGC 2728
817 aPheGluArgGluGlyPheGlyArgArgLysSerMetSerGluLysArgThrL 834
2729 TTTTCAACGAGAGAGATTTGGACGTCAGAGTATGTCAGAAAAAGCCACAA 2778
834 ySglnPheSerAspAlaSerGlnLeuAspPheValLysTyrArgLysSer 850
2779 AGCAATTTTCAGATGCCAGTCAATTTGATTTGCTTAATAAACGAAAAATCA 2828
851 LysSerMetAspLeuGlyIleAlaAspGluThrLysLeuAsnThrValAs 867
2829 AAAAGCATGGAATTAAGTATAGCTGACGACGACATAACTCAATACAGTCCA 2878
867 pAspGluLysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2879 TGACCAAGAAAGCAGATTTCTCCACGACAGAGATGTGGTCTTCCCTGGGCTC 2928
884 euLysLysSerSerSerLeuGluSerLeuGlnThrAlaValAlaGluVal 900
2929 TGAAGAGTCAAGCTCATTTGAGAGTCTGACAGCCGAGTTCGCCGAGGTG 2978
901 ThrLeuAsnGlyAspIleProPheHisArgProArgProArgIleIleArg 917
2979 ACTTTGAATGGGATATATCTTTCATCTGTCACGCGCGCGGATTAATCAG 3028
917 gGlyArgGlyCysAsnGluSerPheArgAlaAlaIleAspLysSerTyrA 934
3029 AGCGAGGGGATGCAATGAGACTTCAGAGCTGCCATTCACAAATCTTAATG 3078
934 sPLysProAlaValAlaAspAspAspGluGlyMetGluThrLeuGluGlu 950
3079 ATAAACCGCGGTAGATGATGATGAAGCATGAGAACCTTGGAAATAA 3128
951 AspThrGluGluSerSerArgSerGlyArgGluSerValSerThrAlaSe 967
3129 GACACAGAAAGAAAGTTCAAGATCAGGGAGAGAGTCTGTATCCACAGCCAG 3178
967 rAspGlnProSerHisSerLeuGluArgGluMetAsnGlyAsnGlnGlu 984
3179 TGATCAGGCTTCCCACTCTCTGGAGACACAAATGAAATGAAACCAAGAGA 3228
984 ySgIlyAspLysThrAspArgLysLysAspLysThrGlyLysGluLysLys 1000
3229 AAGGTGATTAAGACTGATTAAGAAAAAGATTAACGTGAAAAAGAAAGAG 3278
1001 LysAspArgAspLysGluLysAspLysMetLysAlaLysLysGlyMetLe 1017
|||||

```

```

3279 AAAGATAGAGATAGAGAGAGATATAAATGAAAGCCAGAGAGATGCT 3328
1017 uUysgLyLeuGlyAspMetPheArgPheGlyLysHisArgLysAspAspL 1034
3329 GAAGGCGCTTGAGAGACATGCTTACGCTTGCCAACTGAAAGCCGAGAGA 3378
1034 ys 1034
::
3379 GA 3380

seq_name: gb_pr:AF332592

seq_documentation_block:
LOCUS AF332592 2967 bp mRNA linear PRI 16-NOV-2001
DEFINITION Homo sapiens atypical PKC isoform-specific interacting protein
ACCESSION AF332592
VERSION AF332592.1 GI:14579308
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2967)
Fang, C.M. and Xu, Y.H.
TITLE Down-regulated expression of atypical PKC-binding domain deleted
JOURNAL Cell Res. 11 (3), 223-225 (2001)
MEDLINE 21520080
PUBMED 11642408
REFERENCE 2 (bases 1 to 2967)
AUTHORS Fang, C. and Xu, Y.
TITLE Exon/Intron Structure And Splicing Variants of A Novel Human
JOURNAL Polarity Gene, hasip
AUTHORS 3 (bases 1 to 2967)
Fang, C. and Xu, Y.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-2000) Laboratory of Molecular and Cellular
Oncology, Institute of Biochemistry and Cell Biology, Shanghai
Institutes for Biological Sciences, Chinese Academy of Sciences,
310 Yue-Yang Road, Shanghai 200031, P. R. China
LOCATION/Qualifiers
1..2967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10p11.2"
1..2967
/note="ASIPSB: alternatively spliced, similar to GenBank
accession Number AF196186"
/codon_start=1
/product="atypical PKC isoform-specific interacting
protein short variant b"
/protein_id="AAK69192.1"
/db_xref="GI:14579309"
/translacion="MKVTVCFCGRRVVPCGDMVFSIQQAVTRKATAKDPNY
WQVHRLEHGDGGLDDDLCDVADKDIYAVFEDQPHHGGSSTSTGSPSE
IFSEELCTNNVSAPOPYOASTSEVTPSVLRAMPLVRRSSDPLIGISTSVDSNF
SSEPSRKNPTNPTSTTNGTAKONTASPKTCRKDKDENRSLPRTSMNSMVFQDNA
RSLSASHPRVGMLEKQEDQEDGEDENSRVPGVHADHAGLEHLPNLSMDMYTVE
VPDQGPLGIVHVFESARGRTGLLVKRLKLEKGAHEHNLRENDICVRIINDQDLN
RREDOAHMRQAMRTPITWVHYPAANKEOYEOISOSRKNYYSRSFSPSOYIDNR
SVNSAGLHTVORALNHPPEQIDSHSRLPFAHBSCKPSPASAPQNVSTVSSG
YNTKLIKRLNIOLKKTEGIGPISITSRDVTIGGSAPITYKNIIPRGAIDQGRILKAG
DRLIENGVDLVGKQSEYVSLRSTKMGVTSLLVFEQEAHFRELKADEDIYLT
PGDTRELFTEFVINDSGSAGVSVKGRNSKENADIGIPKVISINGGAASKGRAR
VNDOLAVNGESLIGKTNDAMETLRRSKSTEGMRGAIQIIVARISCKEELKSPGS
PPEPEPIETALDDRRERISHSLSYIGLDESPRNALAIIRKGTOLSCWELKSPGS
DYVITIDDRLPVLPPLSDOSSSHSDVGEFTVADAGWAAALISDADCSLSVMPD
VLAFOEGEGRQIADETKLTNYVDQKAGSPSDSGILKSSLSLQTAVAEVL
NDIDPHRRPRRIIGRGCNESFRALIDKSYDKPAVDDDGEMTLEDEESRGR
ESVTSADQPSHSLEROMNGNEKDKTDKDKKDKGKDKKDKKDKKDKKAKKMLK

```

```

BASE COUNT 890 a 653 c 776 g 648 t
ORIGIN
alignment_scores:
Quality: 5014.00 Length: 1034
Ratio: 5.096 Gaps: 3
Percent Similarity: 95.164 Percent Identity: 94.681
alignment_block:
US-09-757-781-2 x AF332592
Align seg 1/1 to: AF332592 from: 1 to: 2967
1 MetLysValThrValAlcysPheGlyArgThrArgValAlaValProCysG1 17
1 ATGAAAGTGACCGTGTGCTCGAGCGACCCGGGTGTGTCGTCGCGCGG 50
17 YAspGlyHisMetLysValPheSerLeuIleGlnGlnAlaValThrArgT 34
51 GCAGCGCCACATGAAAGTTTTCAGCCTCATCCAGCAGCGCGGTGACCGCT 100
34 YArgGlyAlaIleAlaLysAspProAsnTYTTPILegLValHisArg 50
101 ACCGGAAAGCCATCCGCAAGATCCAAACTACTGGATACAGTCATCGC 150
51 LeuGluHisGlyLysPGLYLYTLeuAspLeuAspAspIleLeuCysAs 67
151 TTGGAACATGGAGATGGAGGAAATCTAGACCTTGATACATCTTGTGTA 200
67 pValAlaAspAspLysAspArgLeuValAlaValPheAspGluGlnAsp 84
201 TGTACGACAGCATAAAGACAGACTGTAGCAGTGTGATGAGACAGAGATC 250
84 rOHLShISGLYLYAspGlyTYhrSerAlaSerThrGlyThrGlnSer 100
251 CACATCAGCGAGGTGATGCGACAGTCCAGTTCACGGGTGCCAGAGAC 300
101 ProGluIlePheGlySerGluLeuGlyThrAsnAsnValSerAlaPheG1 117
301 CCAAGATATTTGTAGTGTAGCTTGGCACCAACAAATGTCTCAGCCTTCA 350
117 nProTYGlnAlaThrSerGluIleGluValThrProSerValLeuArg 134
351 GCCTTACCAAGCAACAAGTAATTGAGGTCAACCTTCAGCTTCAGAG 400
134 LAsnMetProLeuHisValArgArgSerSerAspProAlaLeuIleGly 150
401 CAATATGCTCTTCATGTTCAGACGAGTAGACCCAGCTCTAATTGCG 450
151 LeuSerThrSerValSerAspSerAsnPheSerSerGluGluProSerAr 167
451 CTTCTCAGCTTGTGTCAGTATGATTAATTTTCTCTGGAAGGCTTCAAG 500
167 GlyAsnProThrArgTPSerThrAlaGlyPheLeuLysGlnAsnT 184
501 GAAAAATCCACACGCTGTCACAAACAGCTGGCTTCTCAGCAGACAGA 550
184 hPAlaGlySerProLysThrCysAspArgLysLysAspGluAsnTYArg 200
551 CTGCTGGGAGTCTTAAACCTCGCAGCAGGAAGAAAGTGAACACTACAGA 600
201 SerLeuProArgAspThrSerAsnTrpSerAsnGlnPheGlnArgAspAs 217
601 AGCTCCCGCGGGATAGTAACGTGCTAACCACTTTCAGAGAACACA 650
217 nAlaArgSerSerLeuSerAlaSerHisProMetValGlyLysTrpLeuG 234
651 TGCTCGCTCGTCTCTGAGTCCAGTCCAAAGTGCGGCAAGTGGCTGG 700
234 LuysGlnGluGlnAspGluAspGlyThrGluGluAspAsnSerArgVal 250

```

```

701 AGAAACAAGACAGATGAGATGGACAGACAGATTAACACTGCTGTT 750
251 GUpProValGlyHisAlaAspThrGlyLeuGluHisIleProAsnPhse 267
751 GAACCGTTGGACATGCTGACAGGGTTTGGACATTAACCAACTTTC 800
267 rLeuAspAspMetValLysLeuValGluValProAsnAspGlyLeuProL 284
801 TCTGGATGATGTGTAAAGCTCTAGAAAGTCCCAACGATGGAGGCGTC 850
284 enuGlyLeuHisValProPhseSerAlaArgGlyArgThrLeuGly 300
851 TGGGAATCCATGTAGTCCCTTTCAGTCTCGAGCGCGACAGACCTGGGG 900
301 LeuLeuValLysArgLeuGluLysGlyLysAlaGluHisGluAsnLe 317
901 TTATTGTATAAAGCATTTGGAAAGGTGGTAAGCTGAACATGAAATATC 950
317 uPheArgGluAsnAspCysIleValArgIleAsnAspGlyAspLeuArg 334
951 TTTTCGTGAGAAATGATGTGATGTCTGAGATTAATGATGGCGACCTCGAA 1000
334 snArgArgPheGluGluInAlaGluHisMetPheArgGluAlaMetArgThr 350
1001 ATGGAAGATTTGACACAGCACAACTATGTTTGGCCAGGCAATGCGGTACA 1050
351 ProIleIleTrpPheHisValValProAlaAlaAsnLysGluGlnTrpG 367
1051 CCCATCATTTGGTTCCATGTGGTTCCGCGCAAAATTAAGACAGCATATGA 1100
367 uGluLeuSerGlnSerGluLysAsnAsnTrpTyrSerSerArgPheSerP 384
1101 ACACATATCCCAAGAGTGAAGAAACAATTAATTAATCAAGCCGTTTAAAGCC 1150
384 roAspSerGlnTrpIleAspAsnArgSerValAsnSerAlaGlyLeuHis 400
1151 CTGACAGCCGATATGATGACACAGAGGTGTAAACAGTGCAGGCGCTTAC 1200
401 ThrValGlnArgAlaProArgLeuAsnHisProProGluGlnIleAspSe 417
1201 ACGGTGACAGAGACACCCCGACTGAACCCACCCGCTGAGCAGATAGACTC 1250
417 rHisSerArgLeuProHisSerAlaHisProSerGlyLysProProSerA 434
1251 TCACCTAAGACTACTCATGACGACACACCCCTCGGAAACACCAATCCG 1300
434 laProAlaSerAlaProGlnAsnValPheSerThrThrValSerSerGly 450
1301 CTCACGCTCGGCACCTCAGAAATGTATTATGACACTGTAAACAGTGGT 1350
451 TyrAsnThrLysLysIleGlyLysArgLeuAsnIleGlnLeuLysLysG 467
1351 TATTAACACCAAAAATAAGCAGAGAGGCTTAATATCCAGCTTAAGAAAG 1400
467 yThrGluGlyLeuGlyPheSerIleThrSerArgAspValThrIleGlyG 484
1401 TACAGAGAGGTTTGGATTCAGCATCACTTCACAGAGATGTAAACAATAGGT 1450
484 lYserAlaProIleTyrValLysAsnIleLeuProArgGlyAlaAlaIle 500
1451 GCTCAGCTCCAAATCTATGTGAAAAACATTCCTCCCGGGGGGGCGCCATT 1500
501 GlnAspGlyArgLeuLysAlaGlyAspArgLeuIleGluValAsnGlyVal 517
1501 CAGGATGGCGCACTTAAGCAGAGACAGACTTATAGAGTAAATGAGT 1550
517 lAspLeuValGlyLysSerGlnGluGluValValSerLeuLeuArgSerT 534
1551 AGATTTAGTGGCAAAATCCCAAGAGAAATGTTGCTGCTGAGAAAGCA 1600
534 hrLysMetGluGlyThrValSerLeuLeuValPheArgGlnLysAla 550
1601 CCAAGATGGAGAAACTGTGAGCTTCTGCTTTCGCCAGAAAGACGCC 1650

551 PheHisProArgGluLeuAsnAlaGluProSerGlnMetGlnIleProLy 567
1651 TTCACCCCAAGGAACTG..... 1668
567 sGluThrLysAlaGluAspGluAspIleValLeuThrProAspGlyThra 584
1669 .....AAAGCAGAAAGATGAGATATTGTTTACACCTGATGGCACCA 1711
584 rGluPheLeuThrPheGluValProLeuAsnAspSerGlySerAlaGly 600
1712 GGGAAATTTGCAATTTGAAAGTCCCACTTAATGATTCAGGATTCGACGC 1761
601 LeuGlyValSerValLysGlyAsnArgSerLysGluAsnHisAlaAspLe 617
1762 CTGGTGTGCTAGTGTCAAAAGGTAAACGGTCAAAAAGAAACCAACGAGATT 1811
617 uGlyIlePheValLysSerIleIleAsnGlyGlyAlaAlaSerLysAspG 634
1812 GGGAAATCTTTGTCAAAAGTCCATTTAATGAGAGCAGCAGCATCTAAAGATG 1861
634 lYArgLeuArgValAsnAspGlnLeuIleAlaValAsnGlyGluSerLeu 650
1862 GAAGGCTTCGGGTGAATGATCACTGATAGCAGTAAATGGAGAAATCCCTG 1911
651 LeuGlyLysThrAsnGlnAspAlaMetGluThrLeuArgArgSerMetSe 667
1912 TTGGGCAAGAACAAACCAAGATGCCATGGAAACCTTAAGAAAGTATTATGTC 1961
667 rThrGluGlyAsnLysArgGlyMetIleGlnLeuIleValAlaArgArgI 684
1962 TACGAAGGCAATTAACGAGAAATGATCCAGCTTATTTGTTCAAGAGAA 2011
684 lSerLysCysAsnGluLeuLysSerProGlySerProProGlyProGlu 700
2012 TAAGCAAGTGCATGAGCTGAAGTCACTGGAGACCCCTCGACCTGAG 2061
701 LeuProIleGluThrAlaLeuAspAspArgGluArgArgIleSerHisSe 717
2062 CTGGCCATTGAACAGCGTTGGATGATAGAGAAAGAAATTTCCCATTC 2111
717 rLeuTyrSerGlyIleGluGlyLeuAspGluSerProSerArgAsnAlaA 734
2112 CCTTACAGTGGCAATGTAGGGGCTTGATGAATCCGCCAGCAGAAATGCTG 2161
734 lLeuSerArgIleMetGlyGluSerGlyLysTyrGlnLeuSerProThr 750
2162 CCTCAGTAGGATTAATG.....GCTAAATACCAAGCTGTCCCTTACA 2202
751 ValAsnMetProGlnAspAspThrValIleIleGluAspAspArgLeuPr 767
2203 GTGAATATGGCCCAAGATGACACTGTCATTAATAAATATGACAGGTTGCC 2252
767 oValLeuProProHisLeuSerAspGlnSerSerSerSerHisAspA 784
2253 AGTGCCTTCCACATCTCTCTGACAGCTCTTCCAGCAGCTCCCATGATG 2302
784 sPValGlyPheValThrAlaAspAlaGlyThrThrPalLysAlaAlaIle 800
2303 ATGTGGGGTTTGTGACGCGCAGAGTCTGTACTTGGCCAGAGGCTGCATC 2352
801 SerAspSerAlaAspCysSerLeuSerProAspValAspProValAlaI 817
2353 AGTATTCACCGCACTGCTCTTTGAGTCCAGATGTTTATCAGATTCTTGC 2402
817 aPheGlnArgGluGlyPheGlyArgGlnSerMetSerGluLysArgThrL 834
2403 TTTTCAACGAGAAAGATTTGGACGTGAG..... 2430
834 ySGlnPheSerAspAlaSerGlnLeuAspPheValLysThrArgLysSer 850
2430 ..... 2430

```

```

851 lysSerMetaspLeuGlyIleAlaAspGluThrIleuAsnThrValAs 867
2431 .....ATAGCTGACGAGACTAAACTCATATACAGTGA 2462
867 PASpGlnIysAlaGlySerProSerArgAspValGlyProSerLeuGlyL 884
2463 TGACCAGAAACAGGTTCTCCACAGACAGATGTGGTCCCTCCGCTC 2512
884 euLYsLYsSerSerSerLeuGluSerLeuGlnThrAlaValAlaGlu 900
2513 TGAAGAAGTCAAGCTCATGTGAGAGTGTGCAGACCGCATGGCAGAGTG 2562
901 ThrIleuAsnGlyAspIleProPheHisArgProArgIleIleLeu 917
2563 ACTTGAAATGGGAAATTCCTTCATCGTCCAGCGCGGAGTAATACAG 2612
917 gGlyArgGlyCYsAsnGluSerPheArgAlaAlaIleAspLYsSerLYr 934
2613 AGCGAGGGAGATGCAATGACAGCTTCAGAGCTGCCATCGACAAATCTTATG 2662
934 sPLysProAlaValAspAspAspGluGlyMetGluThrLeuGluGlu 950
2663 ATAAACCCGGCGTATGATGATGATGATGATGATGATGATGATGATGAT 2712
951 AspThrGluIleuSerSerArgSerGlyArgGluSerValSerThrAla 967
2713 GACACAGAAAGAAAGTCAAGATCAGGAGAGAGTCTGTATCCACAGCAG 2762
967 rAspGlnProSerHisSerLeuGluArgGlnMetAsnGlyAsnGlnGlu 984
2763 TGATCACCCTTCCTCCACTCTCTGGAGAGACAAATGATGAGAAACCAAG 2812
984 ySGLYsPLysThrAspArgLYsLYsAspLYsThrGlyLYsGluLYsLYs 1000
2813 AAGGTGTAAGACTGATAGAAAAAGATAAACTGAAAAAGAAAAAGAG 2862
1001 LYsAspArgAspLYsGluLYsAspLYsMetLYsAlaLYsLYsGlyMet 1017
2863 AAGATATAGATATAGAGAAAGATATAATTAAGCAAGAGAGGAATGCT 2912
1017 uLYsGlyLeuGlyLYsMetPheArgPheGlyLYsHisArgLYsAspAsp 1034
2913 GAAGGCTTGGGAGACATGTTCAGCTTCACAACTGAAGCCGAGACAGA 2962
1034 ys 1034
2963 GA 2964
seq_name: gb_pat:BD004905
seq_documentation_block:
LOCUS BD004905 2718 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel genes encoding protein kinase or protein phosphatase.
ACCESSION BD004905
VERSION BD004905.1 GI:18632866
KEYWORDS JP 03074935-T/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2718)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Makamatsu,A., Nagai,K., Otsuki,T.,
Funahashi,S., Seno,C. and Nezu,J.
Novel genes encoding protein kinase or protein phosphatase
Patent: JP 03074935-T 2 30-JAN-2001;
TITLE HELIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA,
KOUJI HAYASHI, KAORU SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, OMOYASU
SUGIYAMA, AI WAKAMATSU, KEIICHI NAGAI, TETSUJI OTSUKI, SHINICHI
FUNAHASHI, HIKARI SENO, JUNICHI NEZU
COMMENT OS Homo sapiens (human)
PD JP 03074935-T/2
PD 30-JAN-2001

```

```

PF 28-JUL-2000 JP 2000005061
PR 29-JUL-1999 JP 99P 248036,11-JAN-2000 JP 00P 118776 PR
02-MAY-2000 JP 00P 183767,09-JUN-2000 JP 00P 241899 PR
18-OCT-1999 US 60/159590,17-FEB-2000 US 60/183322 PI TOSHIO
OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,PI
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI
WAKAMATSU,KEIICHI NAGAI,TETSUJI OTSUKI,SHINICHI
FUNAHASHI,HIKARI SENO,
PI JUNICHI NEZU
PC C12N15/12,C12N9/12,C12N9/16,C12N1/15,C12N1/19,C12N1/21,PC
C12N5/10,C07K16/40,
PC C12Q1/68,C12Q1/42,C12Q1/48
CC
FH
FT CDS Location/Qualifiers
1..2718
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 840 a 586 c 688 g 604 t
ORIGIN
FEATURES
Source 1..2718
Location/Qualifiers
1..2718
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 840 a 586 c 688 g 604 t
ORIGIN
Alignment scores:
Quality: 4409.00 Length: 909
Ratio: 5.068 Gaps: 2
Percent Similarity: 95.710 Percent Identity: 94.829
Alignment_block:
US-09-757-781-2 x BD004905 ..
Align seg 1/1 to: BD004905 from: 1 to: 2718
126 GluValThrProSerValLeuArgAlaAsnMetProLeuHisValArg 142
3 GAGTGCACACCTTCAGTCTCGACGAAATATGCTCTTCATGTTCCGCG 52
142 gSerSerAspProAlaLeuIleGlyLeuSerThrSerValSerAspSer 159
53 CAGTAGTACCCAGCTTAATTGGCTCTCCACTTCTGTCTGATGATAGTA 102
159 snPheSerSerGluGluProSerArgLYsAsnProThrArgTrpSerThr 175
103 ATTTTTCCTCGAAGAGCCTTCAGGAAAAATCCACACCGCTGTGCAACA 152
176 ThrAlaGlyPheLeuLYsGlnAsnThrAlaGlySerProLYsThrCYsAs 192
153 ACAGCTGCTCTCCACAGACAGACACTGCTGGAGTCCCTAAAGCTCGCA 202
192 PArgLYsLYsAspGluAsnTrpArgSerLeuProArgAspThrSerAsn 209
203 CAGAGAAAGATGAAACTACAGAACCTCCCGCGGAGTACTAGTAACT 252
209 rPserAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225
253 GGTCTAACCAATTTTCAGAGACAAATCTCGCTCGTCTGAGAGTCCAGT 302
226 HisProMetValGlyLYsTrpLeuGluLYsGlnGluGlnAspGluAsp 242
303 CACCAATGTGGGGCAGAGTGGCAGAGNAGAAACAGACAGATAGAGTGG 352
242 yThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspThr 259
353 GACGAAAGAGATAAACAGTGTGTTGAACTGTGGCATGCTGACACGG 402
259 LYsLeuGluHisIleProAsnPheSerLeuAspAspMetValLYsLeuVal 275
403 GTTTGGAGCATATACCACTTTTCTGTGATATATGTAAGCTCGTA 452
276 GluValProAsnAspGlyLYsProLeuGlyIleHisValAlaProPhe 292
453 GAAGTCCCAACGATGAGAGGCTCTGGGAATTCATGATGCTTTCAG 502
292 rAlaArgGlyGlyArgThrLeuGlyLeuValLYsArgLeuGluLYsG 309

```

```

|||||
503 TGTCTCAGGCGGCAAAACCTGGGGTTATTAGTAAAAAGATTGGAGAAAG 552
309 IYGLIYSLAAGLUHISGLUASNLEUPHEARGLUASNAIPCYSILEVAL 325
553 GTGGTAAGCTGAACATGAAAATCTTTTGTGAGAAATGATTCATGTGTC 602
326 ARGILEASNAAPGLYASPLEUARGASNAARGARGPHEGLUGLINALAGLNH 342
603 AGGATTAAATCATGTGCGCTTCCAAATAGAAAGATTTCAAACACACACACA 652
342 SMETPHEARGSLALAMETARGTHRPROILEIETRPHEHISVALYALP 359
653 TATGTTCCGCAAGCCATGCGTACACCCATCATTTGGTTCATGTGGTTC 702
359 ROALAAASNLYSGLUGLINTYRGLUGLINSERGLINSERGLIULYSASN 375
703 CTCGACGAATAAAGACAGTATGAAACAATATCCCAAAAGTGAGAAAGAC 752
376 ASNTYTYTSERSEARGPHESERPPOASPSEGLINTYRILEASPNANAR 392
753 AATTACTATTCAGCCGCTTTTACGCCGACAGCCAGCATATTTGACACACAG 802
392 GSERVALASERIALAGLYLEUHNSTHRVALGINARGALAPROARGLEUA 409
803 GAGTGTGAACAGTGCAGGGCTTCACACGGTGCAGAGACACCCGACTGA 852
409 SNHISPROFROGLUGLNLLEASPSERHISSEARGLEUPROHISSEARLA 425
853 ACCACCCGCTGAGCAGATGACTCTCACCACAGACTACCTCATAGCGCA 902
426 HISPROSEGLIULYSFROPROSERALAPROALASERIALAPROGLNANVA 442
903 CACCCCTCGGGAAAACCAACCATCGGCTCCAGCCTCGGCACCTCAGATGT 952
442 IPHESERTHTRVALSERSEGLIYTYRASNTHRLYSLILEGLIYLSA 459
953 AATTAGTACACGTGAAGCAGTGGTTATTAACACCAAAAAAATAGGCAAGA 1002
459 RYLEUASNILEGLINLEULYSGLIYTHRGUGLYLEUGLYPHESERIE 475
1003 GGCTTAATATCCAGCTTAAGAAAAGTACAGAAAGTTGGGATTCAGCATC 1052
476 THRSERARGSPVALTHRIEGLIYGLYSERIALPROILETYRVALYLSAS 492
1053 ACTTCCAGAGATGTAAACAATAGTGGCTCAGCTCAATCTATGTGAAAAA 1102
492 NILELEUPROARGLYALALALIEGLINASPGIYARGLEULYSALAGLYA 509
1103 CATTCCTCCCGGGGGGGGGCCATCGAGATGGCCGACTTAAGGCAAGAG 1152
509 SPARGLEULIEGLIUALASNGLYVALASPLEUVALGLIYLSERGLINLU 525
1153 ACAGACTTATAGAGTAAATGGAGTATATTAGTGGCAAAATCCCAAGAG 1202
526 GLUVALVALSERLEUENARGSETRHLYSMETGLUGLYTHRVALSERLE 542
1203 GAAGTGTCTTCGCTGTGAGAACGACCAAGATGGAAGAACTGTGAGACCT 1252
542 ULLEUVALPHEARGSLINLUASPLAPHEHISPROARGLULEUASNLAG 559
1253 TCTGTGCTTTCCGAGGAAGAGCGCTTCCACCCCAAGGGAACGTGATGAG 1302
559 IUPROSEGLINMETGLINILEPROLSGLUHTHRLYSLAAGLUASPGILASP 575
1303 AGCCAAACCGAGATGACATTCCAAAAAGAAAGCAAGCAAGATGAGAGAT 1352
576 ILEVALLEUTHRPROASPGIYTHARGGLUPHELEUTHRPHGLIUALPR 592
1353 ATGCTTCTACCTGATGGCAGCAGGAATTTCTGACATTTGAAGTCC 1402
592 OLEUASNASERGLYSERIALAGLYLEUGLYVALSERVALYSLGLYASNA 609
|||||
1403 ACTTAGTATTCAGGATCTGCAGGCTTGGGTGCAGTGCAAAAGGTAAACC 1452
609 RYSERIYSLUASNHNHISALASPLEUGLYILEPHEVALYSSERILE 625
1453 GGTCAAAAGAGAAACACAGCATTTGGGAATCTTTGTCAACATCCATATAT 1502
626 ASNGLYGLYALIASERLYSASPGIYARGLEUARGVALASNAAPGLINLE 642
1503 AATGAGAGACAGCATCTTAAAGATGGAAGGCTTCGGGTGAATGATWCACT 1552
642 ULLEALVALASNGLYLSERLEULEUGLYLYSTHRASNGLINSPALAM 659
1553 GATAGCAGTAAATGAGAAATCCCTGTTGGGCAAGACAAACCAAGATGCCA 1602
659 ETGLUHTLEUARGARGSEMETSETRHGLUGLYASNLYSARGLYMET 675
1603 TCGAAACCTTAAGAAAGTCTATGCTACTGAGAGCAATTAACGAGAAATG 1652
676 ILEGLINLEUVALALAAARGARGILESERIYSCYASNGIULEULYSE 692
1653 ATCCAGCTTATGTTCGACAGGAGAAATGACAAAGTGAATGAGCTGAAGTC 1702
692 RPROGLISERPROFROGLIYPROGLIULEUPROILEGLUHTRALALEUASPA 709
1703 ACCTGGAGACCCCTGAGCTGAGCTGACCTGCCATTGAAACAGCTTGATG 1752
709 SPARGLUARGARGILESERHISSELEUTYRSEGLIYGLUGLYLEU 725
1753 ATAGCAAGCAAGAAATTTCCCATTCCTCTACAGTGGAGATGAGGGGCTT 1802
726 ASPGLUSERPROSERARGASNAALALALEUSERARGILEMETGLYGLUSE 742
1803 GATGAATCGCCACAGAAAATGCTGCCCTCAGTAGATGAATG..... 1844
742 RGLIYSTYRGLINLEUSERPROTHRVALASMETPROGLINSPASPTHRV 759
1845 .GGTAAATPACAGCTGTCCCTACAGTGAATATATCCCAAGATACACATG 1893
759 ALILEIIEGLUASPASPARGLEUPROVALLEUPROPROHISLEUSERASP 775
1894 TCATTATAGAAAGATGACAGGTGGCAGTCTTCCTCCACATCTCTGTGAC 1943
776 GLINSEUSERSEUSERHISASPSAPVALGLYPHEVALTHRALASPAL 792
1944 CAGTCTCTTCCAGCTCCATGATGATGTGGGTTGTGAGAGCGCAGATGC 1993
792 AGLYTHTRPALYLSAIALALALIESERASPSERIALASPCYSSERLEUS 809
1994 TGTACTTGGCCAAAGGCTGCATATCAGTATTCAGCCGACTGTCTTTGA 2043
809 ERPROASPVALASPPROVALLEUALAPHEGLINARGGLIYPHEGLIYARG 825
2044 GTCCAGATGTTGATCCAGTCTTCTGCTTTCAACGAAAGAAATTTGGAGCT 2093
826 GLINSEMETSERGLIULYSARGTHRLYSGLINPHESEASPALASERGLINLE 842
2094 CAGACT..... 2099
842 UASPPHEVALYSTHRARGLYSERLYSERMETASPLEUGLYILEALAA 859
2100 .....G 2100
859 SPGLUHTHRLYSLUASNTHRVALASPSAGLINSALAGLYSERPROSER 875
2101 ACAGAGACTAAATCAATPACATGTGATGACCAAGAAAGCAGTTCGCCAGC 2150
876 ARGASPVALGLYPROSERLEUGLYLEULYSIYSSERSETRLEUGLUSE 892
2151 AGAGATGGGGTCTTCCCTGGGTCTGAAGAAAGTCAACATCTATTTGGAAG 2200
892 RLEUGLINTHRALVALALAGLUVALTHRLEUASNGLIYSPILLEPROPHEN 909
2201 TCTGCAGACCCGACTGTCCGAGGTGACTTTGAATGGGATATTCTCTTCC 2250

```

909 isargProatrgProatrgilelleargglyArgglyCysasnGluSerphe 925
|||||
2251 ATCTGTCACGGCCGGGATCAATCAGAGGCGATGCAATGAGAGCTTC 2300
926 ATGAlaAlaIleasplysSerTyraSplysProAlaValAspAspAsps 942
|||||
2301 AAGGTGGCCATCGACAATCTTATGATTAACCCGGGTGATGATGATGATA 2350
942 pGluGlyMetGluThrLeuGluGluAspThrGluGluSerSerArgSerg 959
|||||
2351 TGAAGCATGAGGACCTTGAGAAAGACACAGAAAGAAATTCAGATCAG 2400
959 lYargGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGlu 975
|||||
2401 GGAGAGAGCTGTATCCACAGCCAGCATGATCAGCTTCCACTCTCGAG 2450
976 ArgGlnMetAsnGlyAsnGlnGluGlyAspGlyThrAspArgGlySly 992
|||||
2451 AGACAAATGAATGGAACCAAGAAAGGTGATTAAGACTGATGAAAAA 2500
992 sAspLysThrGlyLysGlyLysLysAspArgAspLysGlyLysAspL 1009
|||||
2501 GGATTAACCTGGAAAAAGAAAGAAAGATAGACATTAAGGAGAAAGATA 2550
1009 ysmetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArg 1025
|||||
2551 AAATGAAGCCCAAGAAAGGAATGCTGAAGGCTTGAGACATGTTCAAC 2600
1026 PheGlyLysHisArgLysAspAspLys 1034
|||||
2601 CTTGCCAAACTGAGCCCGGAGAAGA 2627
seq_name: gb_pr:AK027735
seq_documentation_block:
LOCUS AK027735 2718 bp mRNA linear PRI 15-MAY-2001
DEFINITION Homo sapiens cDNA FLJ14829 f1s, clone OVARC100945, moderately
similar to Rattus norvegicus mRNA for atypical PKC specific binding
protein.
ACCESSION AK027735
VERSION AK027735.1 GI:14042633
KEYWORDS oligo capping, f1s (full insert sequence).
SOURCE Homo sapiens ovary, tumor tissue cDNA to mRNA, clone_l1b:OVARC1
clone:OVARC100945.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites) Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuhio,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished
TITLE 2 (bases 1 to 2718)
JOURNAL Isogai,T. and Otsuki,T.
REFERENCE Direct Submission
AUTHORS Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@h1.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction,
5'- & 3'-end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Source Location/Qualifiers
1..2718

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OVARC100945"
/tissue_type="ovary", tumor tissue"
/clone_l1b="OVARC1"
/note="cloning vector: pME18SFL3"
33..2630
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB55330.1"
/translation="MPLHVRSSDPALIGSTSVSDSNESSESRKNPTRWTTAGF
LKQNTAGSPKACDKKDENYRSLPRDTSNNNDQURNASSUSASHPMVKWEKOE
QDDGCEEDNSRVEPVGHADTGLEHITPNFSLDPMVKLVEVPNDGPGIHHVPSARG
GRTGLGLVKRLKGAHEHNLFEHNDICIVRIDGDLRNRFEQAHMFRQAMRTPII
MEHVPANKEOYEOLSOSEKNNYSRSFSDSOYIDNRVNSGLTVORAPLNP
PROIDSHRKLPHSAHPSPKSPASAPASAPVSTVSSGYNTKIKRRLIOLKKEE
GLGFSITSRYVTIGGSAPIYKYNILPFGCAIIOGRKLKAGRLIEVNGVDLVGSQEBY
VSLRSTKMEGTVALVFQEDAFHRELENAESQMIPEYTAEDDILVTPDGTRE
FLTFEVP LSDSGSAGLGVSVKNSRKENHADLGIYKSI INGAASKDGRILRVNDLI
AVNGESLIGRTNODAMETTLRRSMSTEGNKGMIOLIVARRISKNEELKSPGPEL
PIETALDDRERRISHSLYSGIEGLDESPRNALSLRMGYOLSPYNNPQDPVITE
DDRLPVLPPLISDSSSSSHDDVGFVTAADGTAKAAISDADCSLSPDVPYABR
EGFRGQIDETKRLNTVDQKAGSPSRDYGPGLKKSLSLSLOTAYAEVTLNGDIPFH
RPRRIIRGRGNCNPFRAALIDKSTDKPAVDDDCMSTLEDEEESRSRESVTS
DQPSHSLEQMNQENQEKDTRKKDITGKEKKRDKERDKMKAKKMGKGLDMS
LAKIKPEKR"
BASE COUNT 840 a 586 c 688 g 604 t
ORIGIN
alignment_scores:
Quality: 4409.00 Length: 909
Ratio: 5.068 Gaps: 2
Percent Similarity: 95.710 Percent Identity: 94.829
alignment_block:
US-09-757-781-2 x AK027735 ..
Align seg 1/1 to: AK027735 from: 1 to: 2718
126 GluValThrProSerValLeuArgAlaAsnMetProLeuHisValArgAr 142
|||||
3 GAGGTGACACCTTCAGCTCCTTGAGCAAAATATGCCCTTCACATGTTGCGAG 52
142 gSerSerAspProAlaLeuIleGlyLeuSerThrSerValSerAspSera 159
|||||
53 CAGTATGTGACCCAGCTCTAATGGCTCTCCACTTCTGTCATGATGATGTA 102
159 snPheSerSerGluGluProSerArgLysAsnProThrArgTPSerThr 175
|||||
103 ATTTTTCCTCTGAGAGGCTTCAAGAAAAATCCACAGCGTGGTCAACA 152
176 ThrAlaGlyPheLeuLysGlnAsnThrAlaGlySerProLysThrCysAs 192
|||||
153 ACAGCTGGCTTCTCAAGCAGAACACTGCTGGAGTCCCTAAAGCCTGCGA 202
192 PArgLysLysAspGluAsnTyraArgSerLeuProArgAspThrSerAsnT 209
|||||
203 CAGGAGAGAGATGAATAACTACAGAGAGCTCCCGGGGGACTAGTAAGT 252
209 rPserAsnGlnPheGlnArgAspAsnAlaArgSerSerLeuSerAlaSer 225
253 GGTCTAACCAATTTCAGAGAGACATGCTCGCTGCTCTGAGTGCCAGT 302
226 HisPrometValGlyLysTrpLeuGluLysGlnGluGlnAspGluAspG 242
|||||
303 CACCCAAATGGTGGCAGAGCGAGAAACAAAGAAAGAGATGAGAGATGG 352
242 yThrGluGluAspAsnSerArgValGluProValGlyHisAlaAspThrG 259
|||||
353 GACAGAAAGAGATTAACAGTGTGTAACCTGTGACACTGTCGACACGG 402

259 IYLEUGLHUISTIIEPQASNPHESELEUASPMETVALIYSLIEUVAL 275
|||||
403 GTTGGAGCATATACCACTTTCTCTGATGATATGTAAGAGCTTCGTA 452
276 GIUVALPROASNPGLYGIYPROLEUGLYIIEHISVALIAPROPHSE 252
|||||
453 GAAGTCCCAACGATGAGGAGGCTCTGGGAAATCCATGATGCTTCAG 502
292 rAlAaRgGlYlArGThrLeuGlYleuValLYsARgLeuGlYlYsG 309
|||||
503 TGTGAGGCGGAGAACCTTGGGTTATTAGTAAACGATTGGAGAAAG 552
309 lYcLYlYsAlaGlYHISglUaSnLeuPheARgGlUaSnAPcYsIlEVAL 335
553 GTGGTAAACGTGAACATGAAAATCTTTTCGTGAGATGATTCATGTC 602
326 ARGIIeASnAPGLYsPLeuARgASnARgARgPheGlUGlNAlaGlNH 342
603 AGGATTAAATGATGCGACCTTCGAATAGAAATTTGAAACCAAGCAACA 652
342 sMeTPheARgGlNAlaMeLArGThrProIIeIeTrPheHISValIAP 359
|||||
653 TATGTTCCGCAAGCCATGGGTACACCATTCATTGGTTCATGTGTTTC 702
359 rOAlAaSnLYsGlUGlNTrYrGlUGlNLeuSerGlNserGlYlYsASn 375
703 CTGCACCAATAAAGAGCAGTATGAACAACTATCCAAAGTGAAAGAAC 752
376 AaNTYrYrSerSerARgPheSerProASpSerGlNTrYrIleASpASnAR 392
753 AATTACTATTCAGCCGTTTACCTTGACAGCAGCATATATTACAAACAG 802
392 gSERValaASnSERAlaGlYleuHISThrValGlNARgAlaPROARgLeuA 409
803 GAGGTGAACAGTGCAGGGCTTCACAGGGTGAGAGAGCAACCCGACTGA 852
409 snHISProGlUGlNIIeASpSerHISerARgLeuPROHISerAla 425
853 ACCACCCGCTGAGCAGATAGACTTCACCAAGACTACTCTCATACGCCGA 902
426 HISProSerGlYsProProSerAlaPROAlaSERAlaPROGlNAsnVA 442
903 CACCCCTCGGAACACCAACATCGCTCCAGCCTCGCACCTCAGAAATGT 952
442 lPheSerThrThrValSerSerGlYTrAsnThrLYsLYsIIeGlYLYsA 459
953 ATTTAGTACGACTGTAAAGCAGTGTATATAACCAAAAAATAGGACAGA 1002
459 rglEuaSnIIeGlNLeuLYsLYsGlYThrGlUGlYleuGlYpHeSerIIE 475
1003 GGCTTATATCCAGCTTAAGAAAGGTACAGAAAGTTTGGGATTCAGCATC 1052
476 ThrSerARgASpValThrIIeGlYGlYSerAlaPROIleTrYValLYsAS 492
1053 ACTTCCAGAGATGTAAACATAGGTGGCTCAGCTCCAAATCTATGTGAAAA 1102
492 nIleUProARgGlYAlaAlaIIeGlNAspGlYARgLeuLYsAlaGlY 509
1103 CATTTCCCCCGGGGGGGCCCATTCAGATGGCCGACTTAAGCAGAGAG 1152
509 sPARgLeuIIeGlUValASnGlYValASpLeuValGlYLYsSerGlNLU 525
1153 ACAGACTTAAAGGTAAATGAGATGATTAGTGGCAAAATCCCAAGAG 1202
526 GlUValASerLeuLeuARgSerThrLYsMeGlUGlYThrValSerIE 542
1203 GAAGTGTTCCTGCTGTGAGAACACCAAGATGGAAGAACTGTGACCT 1252
542 uleUValPheARgGlNLUASpAlaPheHISProARgGlUleUASnAlaG 559
1253 TCTGCTCTTCGCCAGGAGAGCGCTTCACCAAGGAACTGATATCAG 1302
559 lUProSerGlNMeGlNIIeProLYsGlUThrLYsAlaGlUASpGlUASp 575

|||||
1303 AGCCAAGCCGATGACAGATTCCAAAGAAAGCAAGAGCAAGATGAGAT 1352
576 IIeValLeuThrProASpGlYThrARgGlUPheLeuThrPheGlUValAPr 592
1353 ATTGTTCTTAACACCTGATGAGCACCAGGGAATTTCTGACATTTGAAGTCCC 1402
592 oLeUASnASpSerGlYSerAlaGlYleuGlYIleSerValLYsGlYASnA 609
1403 ACTTAGTATTCAGGATCTGACAGCCCTGGTGTACAGTGTCAAAAGTAAAC 1452
609 rGSerLYsGlUASnHISAlaASpLeuGlYIlePheValLYsSerIIEIIE 625
1453 GGTCAAAAGAGAAACCAAGCAGATTGGGAATCTTTGTCAAGTCCATATT 1502
626 ASnGlYGlYAlaIleSerLYsASpGlYARgLeuARgAlaSnAPGlNLe 642
1503 AATGGAGAGACAGCATCTTAAGATGGAAGGCTTCGGGTGATGATCAACT 1552
642 uIIeAlaValASnGlYlUserLeuEUGlYlYsThrASnGlNAspAlaM 659
1553 GATAGCAGTAAATGAGAAATCCCTGTGGGCAAGACAAACCAAGATGCCA 1602
659 eTcGlUThrLeuARgARgSerMetSerThrGlUGlYASnLYsARgLYMet 675
1603 TGGAAACCCCTAAGACGCTATGCTACTGAAGCAATAAAGCAGCAATG 1652
676 IIeGlNLeuIIeValAlaARgARgIIeSerLYsCYsASnGlUleuLYsSE 692
1653 ATCCAGCTTATGTGCAAGAGAGATTAAGCAAGGCAATAGACTGAAGTC 1702
692 rPROGlYSerPROGlYPROGlUleuPROIleGlUThrAlaLeuASpA 709
1703 ACCTGGAGGCCCCCTGACCTGAGCTGCCCATTTGAACAGCGTTGATG 1752
709 sPARgGlUARgARgIIeSerHISerLeuTrYserGlYIIeGlUGlYleu 725
1753 ATAGAGAACAGAAATTTCCCATTCCTTCACAGTGGGATTTGAGGGCTT 1802
726 ASpGlUserProSerARgASnAlaIleUserARgIIeMeGlYglUSE 742
1803 GATGAATCGCCACAGCAAAATGCTGCTCAGTAGATGAATAGT..... 1844
742 rGlYLYsTrYrGlNLeuSerProThrValaASnMeLProGlNAspASpThrv 759
1845 .GGTAAATACAGCTGTCCTCAGTGAATATGCCCAAGATGACACTG 1893
759 aIIleIIeGlUASpASpARgLeuPROValleuPROHISleuSerASp 775
1894 TCATTATAGAAAGATGACAGGTTGCCAGTGTCTCTCCACATCTCTCGAC 1943
776 GlNserSerSerSerHISAspASpValGlYpHeValThrAlaASpAl 792
1944 CAGTCTCTTCCAGCTCCATGATGATGTGGGTTTGTGAGCGCAGATGC 1993
792 aGlYThrTrPAlaLYsAlaIleSerASpSerAlaASpCYsSerLeus 809
1994 TGGTACTTGGCCCAAGGCTGCAATCACTGATTCAGCCGACTGCTTTGA 2043
809 eRProASpValASpPROValleuAlaPheGlNARgGlUGlYpHeGlYARg 825
2044 GTCACAGATGTGAATCCAGTTCTTGCTTTTCAMACAGAGATTTGGAGCT 2093
826 GlNserMetSerGlYsARgThrLYsGlNpHeSerASpAlaSerGlNLe 842
2094 CAGACT..... 2099
842 uASpPheValLYsThrARgLYsSerLYsSerMetASpLeuGlYIleAlaA 859
2100G 2100
859 sPGLUThrLYsLeuASnThrValASpASpGlNLYsAlaGlYSerProSer 875
|||||

2101 ACGAGACTAACTCAATACAGTGGATGACCAAGAAAGCAGGTTCTCCAGC 2150
876 ArgaspaValGlyProSerLeuGlyLeuLysSerSerSerLeuGluSe 892
2151 AGAGATGTGGGCTCTCCCTGGGTCTGAAGAACTCAAGCTCATGTGAGAG 2200
892 rleuGlnThrAlaValAlaGluValThrLeuAsnGlyAspIleProPhe 909
2201 TCTGCAGACCGCAGTTGCCGAGCTGACTTGAATGGGATATCTCTTCC 2250
909 isArgProArgProArgTyleIleArgGlyArgGlyCysAsnGluSerPhe 925
2251 ATCGCTCCAGCGCCGCGATATCAAGAGCAGGGATGCAATGAGAGCTTC 2300
926 ArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspAspAs 942
2301 AGAGCTGCCATCGCAAAATCTATGATAAACCCGCGTATGATGATGATGA 2350
942 pgluGlyMetGluThrLeuGluGluAspThrGluGluSerArgSerG 959
2351 TGAAGGCGATGAGACCTTGGAGAGACACACAGAAAGTTCAAGATCAG 2400
959 llyArgGluSerValSerThrAlaSerAspGlnProSerHisSerLeuGlu 975
2401 GGAGAGAGCTGTATCCACAGCCAGTATCAGCTTCCACACTCTGGAG 2450
976 ArgGlnMetAsnGlyAsnGlnGluLysGlyAspLysThrAspArgLys 992
2451 AGACAATGAAATGGAACCAAGCAAGAGGTGATAGACTCATGAAAAA 2500
992 sAspLysThrGlyLysGlyLysLysLysAspArgAspLysGluLysAsp 1009
2501 GGATTAATCTGAAAAAGAAAAGAAAGATAGAGATAGAGAGAGATA 2550
1009 ysmetLysAlaLysLysGlyMetLeuLysGlyLeuGlyAspMetPheArg 1025
2551 AATTAAGCCAAAGCAAGAAAGGAAATGCTGAGAGGCTTGGGAGACATGTCAGC 2600
1026 PheGlyLysHisArgLysAspLys 1034
2601 CTTCCTCAACTGAAAGCCGAGAGAGAGA 2627
seq_name: gb_pr:AK024668
seq documentation_block: 2116 bp mRNA linear PRI 29-SEP-2000
LOCUS AK024668 Homo sapiens cDNA: FLJ21015 f1s, clone CAE05730, highly similar to
DEFINITION AF252293 Homo sapiens PAR3 (PAR3) mRNA.
ACCESSION AK024668
VERSION AK024668.1 GI:10437004
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens primary endothelial cells of human coronary artery
CDNA to mRNA, clone_id:CAE clone:CAE05730.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isegai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2116)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isegai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center,
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
E-mail:cdna@aims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of

FEATURES
source International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
location/Qualifiers
1..2116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="primary endothelial cells of human coronary
artery"
/clone="CAE05730"
/clone_id="CAE"
/note="cloning vector pME18SFL3"
misc-feature 1..2116
/note="highly similar to AF252293 Homo sapiens PAR3 (PAR3)
mRNA"
BASE COUNT 651 a 476 c 571 g 418 t
ORIGIN
alignment_scores: Length: 683
Quality: 3512.50 Gaps: 1
Ratio: 5.165
Percent Similarity: 99.561 Percent Identity: 99.561
alignment_block:
US-09-757-781-2 x AK024668 ..
Align seg 1/1 to: AK024668 from: 1 to: 2116
674 GlyMetIleGlnLeuIleValAlaArgArgIleSerLysCysAsnGlu 690
|||||
2 GGAATGATCCACGCTTATGTTGCAAGAGAGAAATAGCAAGTCAGATGAGCT 51
690 ulysSerProGlySerProProGlyProGluLeuProIleGluThrAla 707
|||||
52 GAGTCACCTGGAGCCGCCCTGGACCTGAGCTGACCTCATGAAACAGCCT 101
707 euAspAspArgGluArgArgIleSerHisSerLeuLysSerGlyIleGlu 723
|||||
102 TCGATGATAGAGAGAACGAAATTTCCCTCTCTACAGGGGATGAG 151
724 GlyLeuAspGluSerProSerArgAsnAlaAlaLeuSerArgIleMetG 740
|||||
152 GGGCTTGATGATCCGCCAGCAGAAATGCTGCCCTCAGTAGATTAATG. 199
740 ygluSerGlyLysTyrGlnLeuSerProThrValAsnMetProGlnAsp 757
|||||
200GGTAATATACACAGCTGCTCCCTACAGTGAATATGCCCAAGATG 242
757 spthrValIleIleGluAspAspArgLeuProValLeuProHisLeu 773
|||||
243 ACAGTGTATTAAGAAATGACAGAGTTGCCAGTCTCTCCACATCTC 292
774 SerAspLysSerSerSerSerHisAspAspValGlyPheValThrAl 790
|||||
293 TCTGACAGCTCTCTCCACCTCCATGATGATGGGTTGTGACGGC 342
790 aAspAlaGlyThrTyrAlaLysAlaIleSerAspSerAlaAspCys 807
|||||
343 AGATGCTGCTGTTGGGCCAAGCTCAATCACTGATTCACCGACTGCT 392
807 erLysSerProAspValAspProValLeuAlaPheGlnArgGluGlyPhe 823
|||||
393 CTGTGAGTCCAGATGTTGATTCAGTTCTTTCATACAGAGAGATTT 442
824 GlyArgGlnSerMetSerGluLysArgTyrThrGlnPheSerAspAla 840
|||||
443 GGACCTCAGAGTATCTCGAAAAACGCAAGCAATTTTCAGATGCGAG 492
840 rGlnLeuAspPheValLysThrArgLysSerLysSerMetAspLeuGly 857

```
|||||
493 TCATGATGATTGCTTAACACGAAATCAAAAAGCATGATTAGGTA 542
857 lealaaspgluThrLysLeuasnThrValAspAspGlnLysAlaGlySer 873
|||||
543 TAGGTGACGAGACTAACTCAATACAGTGCAGAGCAAGGAGTTCT 592
874 ProSerArgAspValGlyProSerLeuGlyLeuLysSerSerSerLe 890
|||||
593 CCAGCAGAGATGTGGTCTCTCCCTGGGTCTGAAGAACTCAAGCTGTT 642
890 uGluSerLeuGlnThrAlaValAlaGluValThrLeuasnGlyAspIleP 907
|||||
643 GGAGAGCTCGACACCGCAGTGCAGGTGACTTGAATGGGATATTC 692
907 roPheHisArgProArgProArgIleIleArgGlyArgGlyCysAsnGlu 923
|||||
693 CTTTCCATCGTCCCGCCGCGGATATCAGAGCAGGGATGCAATGAG 742
924 SerPheArgAlaAlaIleAspLysSerTyrAspLysProAlaValAspAs 940
|||||
743 AGCTTCAGACTGCCATCGACAAATCTTATGATTAACCCGCGGTAGATGA 792
940 pAspAspGluGlyMetGluThrLeuGluGluAspThrGluGluSerSerA 957
|||||
793 TGATGATGAAGGCAATGAGACCTTGAAAGAGACAGAAAGAAAGTTCAA 842
957 rGSerGlyArgGluSerValSerThrAlaSerAspGlnProSerHisSer 973
|||||
843 GATCAGGGAGAGAGCTCTGATCCACAGCAGCTGATCAGCCTCCACTCT 892
974 LeuGluArgGlnMetAsnGlyAsnGlnGluLysGlyAspLysThrAspAr 990
|||||
893 CTGGAGAGACAAATGAATGCAACCAAGAGAAAGATAGATAGAGTAG 942
990 glyLysAspLysThrGlyLysGlyLysLysLysAspArgAspLysGlu 1007
|||||
943 AAAAAGAGATAAACTGAAAAAGAAAGAAAGAAAGATAGATAGAGGA 992
1007 ysaAspLysMetLysAlaLysLysLysMetLeuLysGlyLeuGlyAspMet 1023
|||||
993 AGGATAAATATCAAAAGCCAGAGAGGATGCTGAAGGCTTGAGACACTG 1042
1024 PheArgPheGlyLysHisArgLysAspAspLysIleGluLysThrGlyLy 1040
|||||
1043 TTCAGGTTTGGCAAAACATCGAAAGATGACAAAGATTGAGAAACGGGTA 1092
1040 sIleLysIleGlnGluSerPheThrSerGluGluGluArgIleArgMetL 1057
|||||
1093 AATAAAAATACAGGAATCTTTACATCAGAGAGAGAGATACGAATGA 1142
1057 ysgGlnGlnGlnLuarGileGlnAlaLysThrArgGluPheArgGluArg 1073
|||||
1143 AGCAGAGACAGAGAGATTCAGACCAAACTCGAAGATTTAGGAAACGA 1192
1074 GlnAlaArgGluArgAspTyrAlaGluIleGlnAspPheHisArgThrPh 1090
|||||
1193 CAAGCTCGAGAGCCTGACTATGCTGAATTCAGAAATTTTCATCGACATT 1242
1090 eGlyCysAspAspGluLeuMetTyrGlyGlyValSerSerTyrGluGlys 1107
|||||
1243 TGGCTGTGATGATGAGTTAATGTATGGGGAGTTCTTCTTATGAAAGGTT 1292
1107 eRmetAlaLeuAsnAlaArgProGlnSerProArgGluGluHisMetMet 1123
|||||
1293 CCATGGGCTCTCAACGCTAGACCTCAGAGCCACAGAAAGGCGCATATGATG 1342
1124 AspAlaLeuTyrAlaGlnValLysLysProArgAsnSerLysProSerPr 1140
|||||
1343 GATGCTTTGTATGCCCAAGTCAAGAAAGCCCGGAATTCAAAACCTTCACC 1392
1140 oValAspSerAsnArgSerThrProSerAsnHisAspArgIleGlnArgL 1157
|||||
1393 TGTAGACAGTAACAGATCACTCTAGCAATCATGATCGGATACAGCGTC 1442
1157 euArgGlnGluPheGlnGlnAlaLysGlnAspGluAspValGluAspArg 1173
|||||
1443 TGAGGCAAGAAATTTCAAGCAACCAAGCAAGATGAAGATGTGAAGATCGT 1492
1174 ArgArgThrTyrSerPheGluGlnProThrProAsnAlaArgProAlaTh 1190
|||||
1493 CGCGGAGACTATAGTTTGTAGCAAAACCTCGGCCGAAGCAGCGCGCGGAC 1542
1190 rGlnSerGlyArgHisSerValSerValGluValGlnMetGlnArgGlnAla 1207
|||||
1543 GCAGAGCGGCGACACTGTGTCTCGTGGAGGTGCAATGACGCGGAC 1592
1207 rGlnGlnGluArgGluSerSerGlnGlnAlaGlnArgGlnTyrSerSer 1223
|||||
1593 GCGAGAGAGAGACCGGAGAGCTCCAGAGCGGCGCCAGCGCATACAGCTCT 1642
1224 LeuProArgGlnSerArgLysAsnAlaSerSerValSerGlnAspSerTr 1240
|||||
1643 CTGCTCGGCAAGCAGGAAAAATGCCAGCTCGGTCTCCAGGACTCTTG 1692
1240 pGluGlnAspTyrSerProGlyGluGlyPheGlnSerAlaLysGluAsnP 1257
|||||
1693 GGAGCAGAACTACTCCCTCGGGGAAAGCTTCCAGAGTCCAAAGAAACC 1742
1257 roArgTyrSerSerTyrGlnGlySerArgAsnGlyTyrLeuGlyGlyHis 1273
|||||
1743 CCAGGTACTCCAGCTACCAAGGCTTCAGAAACGCTACTACGAGAGACAT 1792
1274 GlyPheAsnAlaArgValMetLeuGluThrGlnGluLeuArgGlnGln 1290
|||||
1793 GCGTTCAACGCCAGGGCTCATCTGGAATCTCAGAGCTCTCCGCCACGA 1842
1290 uGlnArgArgLysGlnGlnGlnMetLysLysGlnProProSerGluGlyP 1307
|||||
1843 ACAGAGCGCGGAAGGAGCAGCAGATGAAGAGCAGAGCTCTCCGAGGGGC 1892
1307 roSerAsnTyrAspSerTyrLysLysValGlnAspProSerTyrAlaPro 1323
|||||
1893 CCAGCAACTATACACTCGTATTAAGAAAGTCCAGACCCACAGTACGCCCT 1942
1324 ProLysGlyProPheArgGlnAspValProProSerProSerGlnValAla 1340
|||||
1943 CCCAAGGGGCCCTTCCGCCAAGATGTGCCCCCTCCCTTCACAGGTTCG 1992
1340 aArgLeuAsnArgLeuGlnThrProGluLysGlyArgProPheTyrSer 1356
|||||
1993 GAGCTGAACAGACTTCAGACTCTCGAAGAAAGGAGGCCCTTCTATTTCC 2041
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:29:09 ; Search time 66.93 Seconds
(without alignments)
3504.873 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 7035
Sequence: 1 MKYVCFGRTRVVPCGDGH.....SQVARNRLQTPKGRPFYS 1356

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7008.5	99.6	1353	4	Q9BY58
2	6537.5	92.9	1273	4	Q96RM6
3	6485	92.2	1266	4	Q9NYE5
4	6332.5	90.0	1337	11	Q9Z340
5	6310.5	89.7	1333	11	Q99NM2
6	5268.5	74.9	1031	4	Q9BY57
7	5014	71.3	988	4	Q96RM7
8	4361	62.0	865	4	Q96K28
9	3335	47.4	667	4	Q9HC48
10	2061.5	29.3	1205	4	Q96NM7
11	1825.5	25.9	1143	4	Q96NM6
12	1822	25.9	347	4	Q9NWL4
13	1198	17.0	624	4	Q96DK9
14	1154.5	16.4	1464	5	Q9VX75
15	1149.5	16.3	1464	5	Q96782
16	1110.5	15.8	545	4	Q96N09

17	628.5	8.9	1376	5	Q95Q69
18	624	8.9	1379	5	Q17353
19	396.5	5.6	213	11	Q9CSB4
20	362.5	5.2	2042	4	Q75970
21	361.5	5.1	2054	11	Q55164
22	356	5.1	1756	5	Q9YBE4
23	352.5	5.0	2055	11	Q9ZIK3
24	346.5	4.9	2208	5	Q09515
25	338	4.8	1134	4	Q9H3N9
26	337.5	4.8	1462	4	Q96Q27
27	337.5	4.8	1524	4	Q15249
28	337.5	4.8	1552	4	Q60833
29	337.5	4.8	1582	4	Q43742
30	329.5	4.7	612	11	Q70471
31	326	4.6	2484	6	Q28006
32	323	4.6	967	5	Q9H179
33	323	4.6	1064	5	Q18165
34	322.5	4.6	2460	11	Q64512
35	319.5	4.5	2766	11	Q9QZ88
36	311	4.4	453	4	Q43798
37	308	4.4	1427	4	Q9NU94
38	303	4.3	526	11	Q08783
39	298	4.2	346	11	Q9CS43
40	297.5	4.2	1492	4	Q9NU93
41	296.5	4.2	5327	5	Q76891
42	289.5	4.1	1410	4	Q9NSN7
43	288.5	4.1	721	11	Q91WJ1
44	286.5	4.1	835	4	Q96C69
45	286.5	4.1	1564	4	Q14160

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	1353 AA.
Q9BY58	Q9BY58	01-JUN-2001 (TREMBLrel, 17, Created)		
AC	Q9BY58	01-JUN-2001 (TREMBLrel, 17, last sequence update)		
DT	01-JUN-2001 (TREMBLrel, 17, last sequence update)			
DT	01-DEC-2001 (TREMBLrel, 19, last annotation update)			
DE	ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN LONG VARIANT.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Fang C., Xu Y.;			
RT	"Exon/Intron Structure and Splicing Variants of a Novel Human Polarity			
RT	Gene, hASIP."			
RL	Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF196185; AAK27891.1; -			
DR	HSSP: Q12923; 3PDZ.			
DR	InterPro: IPR001478; PDZ.			
DR	Pfam: PF00595; PDZ, 3.			
DR	SMART: SM00228; PDZ, 3.			
DR	PROSITE: P550106; PDZ, 3.			
SO	SEQUENCE 1353 AA; 151149 MW; BC531577B9C31AE3 CRC64;			

Query Match 99.6%; Score 7008.5; DB 4; Length 1353;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1353; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY	1	MKYVCFGRTRVVPCGDGHMKVFSLIQAVTRRRRAIAKDPYMWIQVHLEHGSGILD 60
DB	1	MKYVCFGRTRVVPCGDGHMKVFSLIQAVTRRRRAIAKDPYMWIQVHLEHGSGILD 60
QY	61	LDDIICDVADDDKRLVAVFDEDDPHHGGDSSTGTGSPETFGSELGNNVASAPQPYQ 120
DB	61	LDDIICDVADDDKRLVAVFDEDDPHHGGDSSTGTGSPETFGSELGNNVASAPQPYQ 120

```

QY 121 ATSEIETVPSVLRAANPLHVRSSDPALIGLSTSVSDSNFSESPSRKNPTWSTTAGFL 180
Db 121 ATSEIETVPSVLRAANPLHVRSSDPALIGLSTSVSDSNFSESPSRKNPTWSTTAGFL 180
QY 181 KONTGSPKTCORCKKDEYRSPLPRTSNWSNOFORDNARSSLASHPVWGKLEKOEODE 240
Db 181 KONTGSPKTCORCKKDEYRSPLPRTSNWSNOFORDNARSSLASHPVWGKLEKOEODE 240
QY 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVVKLVEVPNDGSGITGVPPSANGRTLG 300
Db 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVVKLVEVPNDGSGITGVPPSANGRTLG 300
QY 301 LLVRLKEGKGAHEHNLFRENDCTVRINDGDLRNRREFQAOHMFROAMRPTIIMFHVPA 360
Db 301 LLVRLKEGKGAHEHNLFRENDCTVRINDGDLRNRREFQAOHMFROAMRPTIIMFHVPA 360
QY 361 ANKEOYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLTVQAPRLNHPPOIDSHSR 420
Db 361 ANKEOYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLTVQAPRLNHPPOIDSHSR 420
QY 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGSITSRDY 480
Db 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGSITSRDY 480
QY 481 TIGGSAPITVKNILPRGAALODGRLKAGDRLIEVNGVDYLGKSOEYVSLLNSTKMEGY 540
Db 481 TIGGSAPITVKNILPRGAALODGRLKAGDRLIEVNGVDYLGKSOEYVSLLNSTKMEGY 540
QY 541 SLVYFROEDAFHRELNAEPSONQIPRETKAEDIEDVILTPDGRREFLTFEVPINDSGSAG 600
Db 541 SLVYFROEDAFHRELNAEPSONQIPRETKAEDIEDVILTPDGRREFLTFEVPINDSGSAG 600
QY 601 LGVSVAGNRSKEHNDLGFVKSIIINGAASDGRRLAVNDOLIANVGESLIGKTODAME 660
Db 601 LGVSVAGNRSKEHNDLGFVKSIIINGAASDGRRLAVNDOLIANVGESLIGKTODAME 660
QY 661 TLRSNSTGKNKGMQILIVARRISKCNELKSPSPPELDELIALDDRERISHSLYS 720
Db 661 TLRSNSTGKNKGMQILIVARRISKCNELKSPSPPELDELIALDDRERISHSLYS 720
QY 721 GIEGLDESRRNALSRINGESGKYOLSPVTMMPODVTIIEDDRLPYLPHLSDQSSSS 780
Db 721 GIEGLDESRRNALSRINGESGKYOLSPVTMMPODVTIIEDDRLPYLPHLSDQSSSS 780
QY 781 SHDDVGYTADAGTAKAAISDSADCSLSPVDVPLAQREGFGROSSEKRTQFSDAS 840
Db 781 SHDDVGYTADAGTAKAAISDSADCSLSPVDVPLAQREGFGROSSEKRTQFSDAS 840
QY 841 OLDFKTRKSKSMOGLIADETKLTNDVQKAGSPSRDVGPSLGLKSSLSLQTAVAEV 900
Db 841 OLDFKTRKSKSMOGLIADETKLTNDVQKAGSPSRDVGPSLGLKSSLSLQTAVAEV 900
QY 901 TLNGDIPFHRPRPRLIRGCGNESFRAIDKSYDKPAVDDDEGEMETLEEDTESSSSGR 960
Db 901 TLNGDIPFHRPRPRLIRGCGNESFRAIDKSYDKPAVDDDEGEMETLEEDTESSSSGR 960
QY 961 ESVAASPOPSHSLEROMNGNOEGKDTOKKDKTGKPKKDRKEDKMAKAKGMLKGL 1020
Db 961 ESVAASPOPSHSLEROMNGNOEGKDTOKKDKTGKPKKDRKEDKMAKAKGMLKGL 1020
QY 1021 GDMERFGKHRRDKLETKGKIKIOESFTSEEBRIKMOBOBERIOAKTREFEREQARERDY 1080
Db 1021 GDMERFGKHRRDKLETKGKIKIOESFTSEEBRIKMOBOBERIOAKTREFEREQARERDY 1080
QY 1081 AEIODEFRTGCGDELMYGVSSYEGSMALNARPOSREGHMADALYAQVKKRNKPSPP 1140
Db 1081 AEIODEFRTGCGDELMYGVSSYEGSMALNARPOSREGHMADALYAQVKKRNKPSPP 1140
QY 1141 VDSNRSTPSNHDRIQRLROFOAKODEDEVEDBRRTYSEQPPNAPRPTOSGRHSVAYE 1200
Db 1141 VDSNRSTPSNHDRIQRLROFOAKODEDEVEDBRRTYSEQPPNAPRPTOSGRHSVAYE 1200
QY 1201 VQOROROEERESSQAOAROYSSLPQSRKNASVSQDSMEQONVSPGEGFQSAKENPRYS 1260

```

```

Db 1198 VQOROROEERESSQAOAROYSSLPQSRKNASVSQDSMEQONVSPGEGFQSAKENPRYS 1257
QY 1261 SYGSRNGITGIGENARVWLETQELLROORRKEQOMKOPPSGSPSNYSDYKKRVOPPS 1320
Db 1258 SYGSRNGITGIGENARVWLETQELLROORRKEQOMKOPPSGSPSNYSDYKKRVOPPS 1317
QY 1321 YAPPKGPFRODVPSPSOVARLNRLQTPPEKGRPPYS 1356
Db 1318 YAPPKGPFRODVPSPSOVARLNRLQTPPEKGRPPYS 1353

RESULT 2
ID Q96RM6 PRELIMINARY; PRT: 1273 AA.
AC Q96RM6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DI 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ATYPICAL PKC ISOYTYPE-SPECIFIC INTERACTING PROTEIN LONG VARIANT B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fang C., Xu Y.;
RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity Gene, hASIP.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF332593; AAK69193.1; -
SQ SEQUENCE 1273 AA; 141730 MW; 53C1A94D8CB7341E CRC64;

Query Match 92.9%; Score 6537.5; DB 4; Length 1273;
Best Local Similarity 93.9%; Pred. No. 0;
Matches 1273; Conservative 0; Mismatches 0; Indels 83; Gaps 4;

QY 1 MKTVYCGRTRRVYVPCGDMKMFSLIOQAVTRYRAIAKDPYWIQVHRLHGDGILD 60
Db 1 MKTVYCGRTRRVYVPCGDMKMFSLIOQAVTRYRAIAKDPYWIQVHRLHGDGILD 60
QY 61 LDDILCDVADDKRLVAVDEODPHHGCGDTSASSGTQSPETFGSELCTNNVSAFOPIQ 120
Db 61 LDDILCDVADDKRLVAVDEODPHHGCGDTSASSGTQSPETFGSELCTNNVSAFOPIQ 120
QY 121 ATSEIETVPSVLRAANPLHVRSSDPALIGLSTSVSDSNFSESPSRKNPTWSTTAGFL 180
Db 121 ATSEIETVPSVLRAANPLHVRSSDPALIGLSTSVSDSNFSESPSRKNPTWSTTAGFL 180
QY 181 KONTAGSPKTCORCKKDEYRSPLPRTSNWSNOFORDNARSSLASHPVWGKLEKOEODE 240
Db 181 KONTAGSPKTCORCKKDEYRSPLPRTSNWSNOFORDNARSSLASHPVWGKLEKOEODE 240
QY 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVVKLVEVPNDGSGITGVPPSANGRTLG 300
Db 241 DGTEDNSRVEPVGHADTGLEHIFPNFSLDDVVKLVEVPNDGSGITGVPPSANGRTLG 300
QY 301 LLVRLKEGKGAHEHNLFRENDCTVRINDGDLRNRREFQAOHMFROAMRPTIIMFHVPA 360
Db 301 LLVRLKEGKGAHEHNLFRENDCTVRINDGDLRNRREFQAOHMFROAMRPTIIMFHVPA 360
QY 361 ANKEOYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLTVQAPRLNHPPOIDSHSR 420
Db 361 ANKEOYEOSSEKKNYSSRFPSPSOYIDNRSVNSAGLTVQAPRLNHPPOIDSHSR 420
QY 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGSITSRDY 480
Db 421 LPHSAHSGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLKKTGEGIGSITSRDY 480
QY 481 TIGGSAPITVKNILPRGAALODGRLKAGDRLIEVNGVDYLGKSOEYVSLLNSTKMEGY 540

```

Db 481 TTGGAPITVYKNIILPRGAIIODGRLKADRLLEVNGVULVCKSOEYVSLRSTMECTV 540
QY 541 SLVFRQEDAFHPRELNAEPESQOMQIPKETADEDEDIVLPDGTREPLTFEYPLNDGSAG 600
Db 541 SLVFRQEDAFHPREL-----KADEDEDIVLPDGTREPLTFEYPLNDGSAG 587
QY 601 LGVSATGKRSKKNHADLGIIFVKSITINGAASKDGLRVNDOLIANVGSLLGKTQNDAME 660
Db 588 LGVSATGKRSKKNHADLGIIFVKSITINGAASKDGLRVNDOLIANVGSLLGKTQNDAME 647
QY 661 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHSLYS 720
Db 648 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHSLYS 707
QY 721 GIEGLDESPSRNAALSRIINGESGKYQLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 780
Db 708 GIEGLDESPSRNAALSRIIM---GKYQLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 764
QY 781 SHDVGFTYADAGTWAKAAISDSADCSLSPVDVPLAFQREBGFRO----- 840
Db 765 SHDVGFTYADAGTWAKAAISDSADCSLSPVDVPLAFQREBGFRO----- 810
QY 841 QLDFVTKRSKSMGLIDETKRLNTVDDQKAGSPSRDYGPSLGLKSSSLESLOTAVAEV 900
Db 811 -----IADETKRLNTVDDQKAGSPSRDYGPSLGLKSSSLESLOTAVAEV 854
QY 901 TLNGDIPHRPRRIIRGRCNESFRRAIDKSYDKPAVDDEDEMETLEEDTEESSRSGR 960
Db 855 TLNGDIPHRPRRIIRGRCNESFRRAIDKSYDKPAVDDEDEMETLEEDTEESSRSGR 914
QY 961 ESVSATSQPQSHSLERONNGQEKDTRDKDKTKGKKKADROKMKAKKMKLGL 1020
Db 915 ESVSATSQPQSHSLERONNGQEKDTRDKDKTKGKKKADROKMKAKKMKLGL 974
QY 1021 GDMFFGKHRRDKLETKGIKIKIQESFTSEERIIMKQEOERIQAKTRFEFRQARNDY 1080
Db 975 GDMF-----RIQAKTRFEFRQARNDY 997
QY 1081 AEIODFHRTFCGDDLELMTGVSSYEGSMALNARPOSREGHMDALYAQVKPRNSKPS 1140
Db 998 AEIODFHRTFCGDDLELMTGVSSYEGSMALNARPOSREGHMDALYAQVKPRNSKPS 1057
QY 1141 VDSNSTPSSNDRIQRLKQEQAKODDEDVEDRRRTYSFEQWPNARPATQSGHSYSVE 1200
Db 1058 VDSNSTPSSNDRIQRLKQEQAKODDEDVEDRRRTYSFEQWPNARPATQSGHSYSVE 1117
QY 1201 VQMQOROEERESSQOARQYSSILPROSKKNASYSODSWEQNTSPGEGFQSAKENPRYS 1260
Db 1118 VQMQOROEERESSQOARQYSSILPROSKKNASYSODSWEQNTSPGEGFQSAKENPRYS 1177
QY 1261 SYQSRNGYLGHGFGNANVLETOELNROQRREQOMKQPPSEGPSNYDSYKRVQDS 1320
Db 1178 SYQSRNGYLGHGFGNANVLETOELNROQRREQOMKQPPSEGPSNYDSYKRVQDS 1237
QY 1321 YAPKGPFRQDVPSPSOVARLNRLQTEPKGRPEYS 1356
Db 1238 YAPKGPFRQDVPSPSOVARLNRLQTEPKGRPEYS 1273

RESULT 3

ID 09NYE6 PRELIMINARY: PRT: 1266 AA.
AC 09NYE6: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PAR3.
GN PAR3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE=20394296; PubMed=10934474;
RA Jobery G., Petersen C., Gao L., Macara I.G.;
RT "The cell-polarity protein Par6 links Par3 and atypical protein kinase
C to Cdc42";
RL Nat. Cell Biol. 2:531-539(2000).
DR EMBL: AF252293; AAF71530.1; -;
DR HSSP: Q12923; 3PD2.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PS0106; PDZ; 3.
SQ SEQUENCE 1266 AA; 141071 MW; 92DE51B68081A4A2 CRC64;

Query Match 92.2%; Score 6485; DB 4; Length 1266;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 90; Gaps 4;

QY 1 MKVTVCGRTRVYVPCGDGMMKVFSLIQQAVTRYKAIKADPNYWIQVHLEHGDGILD 60
Db 1 MKVTVCGRTRVYVPCGDGMMKVFSLIQQAVTRYKAIKADPNYWIQVHLEHGDGILD 60
QY 61 LDDIICDVAADKDRLYAVFDEDDPHHGCDGTSASTGTQSPETFGSELCTNNVSAFQPYQ 120
Db 61 LDDIICDVAADKDRLYAVFDEDDPHHGCDGTSASTGTQSPETFGSELCTNNVSAFQPYQ 120
QY 121 ATSEIEVPSVLRANMPLVHRSSDPALIGLSTVSQDSNFSSEEPERKPTMTSTAGFL 180
Db 121 ATSEIEVPSVLRANMPLVHRSSDPALIGLSTVSQDSNFSSEEPERKPTMTSTAGFL 180
QY 181 KONTAGSPKTCDDKKNDENYRSLPRDTSNWSNOFORNANSSLSASHPMGKMLEKODE 240
Db 181 KONTAGSPKTCDDKKNDENYRSLPRDTSNWSNOFORNANSSLSASHPMGKMLEKODE 240
QY 241 DGTEDNSRVEYVGAADTGLEHINPNFSLDDMKLYAVPNDGCPGLIHVPFSARGRTLG 300
Db 197 DGTEDNSRVEYVGAADTGLEHINPNFSLDDMKLYAVPNDGCPGLIHVPFSARGRTLG 256
QY 301 LLYKRLKGGKKAHEHNLFRFENDCIYRINDGDLRNRFEOAQMHRAMRTPIIMFHVPA 360
Db 257 LLYKRLKGGKKAHEHNLFRFENDCIYRINDGDLRNRFEOAQMHRAMRTPIIMFHVPA 316
QY 361 ANKEOYEOLSOSEKKNYYSRSPSQYIDNRSVNSAGLHTVQARLNNHPEQIDSHSR 420
Db 317 ANKEOYEOLSOSEKKNYYSRSPSQYIDNRSVNSAGLHTVQARLNNHPEQIDSHSR 376
QY 421 LPHSAHPGSKPPSAPASAPQNFSTTVSSGYNTKTIQKRLNIOLKKGTEGLGFSITSRDV 480
Db 377 LPHSAHPGSKPPSAPASAPQNFSTTVSSGYNTKTIQKRLNIOLKKGTEGLGFSITSRDV 436
QY 481 TTGGSAPITVYKNIILPRGAIIODGRLKAGDRLLEVNGVULVCKSOEYVSLRSTMECTV 540
Db 437 TTGGSAPITVYKNIILPRGAIIODGRLKAGDRLLEVNGVULVCKSOEYVSLRSTMECTV 496
QY 541 SLVFRQEDAFHPRELNAEPESQOMQIPKETADEDEDIVLPDGTREPLTFEYPLNDGSAG 600
Db 497 SLVFRQEDAFHPREL-----KADEDEDIVLPDGTREPLTFEYPLNDGSAG 543
QY 601 LGVSATGKRSKKNHADLGIIFVKSITINGAASKDGLRVNDOLIANVGSLLGKTQNDAME 660
Db 544 LGVSATGKRSKKNHADLGIIFVKSITINGAASKDGLRVNDOLIANVGSLLGKTQNDAME 603
QY 661 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHSLYS 720
Db 604 TLRRSMSTEGNKGMIQILIVARRISKCNELKSPSGPELPIETALDRERRISHSLYS 663
QY 721 GIEGLDESPSRNAALSRIINGESGKYQLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 780
Db 664 GIEGLDESPSRNAALSRIIM---GKYQLSPTVNMPODDTVIIEEDRLPYLPPLHSDQSSSS 720
QY 781 SHDVGFTYADAGTWAKAAISDSADCSLSPVDVPLAFQREBGFROSGMSKRTKQPSDAS 840

QY 1081 AEIODEFHRTFGCDDDELMATGVSSYEGSMALNARPOS PREGHMMDALYAQVKRRNSKPS 1140
DB 1081 AEIODEFHRTFGCDDDELMATGVSSYEGSMALNARPOS PREGHMMDALYAQVKRRNSKPS 1139
QY 1141 VDSNRTSTSNHNRIRQLRQEFQAKODEVEYRRRTTSEQOWPNARPAQYSGRHSVSE 1200
DB 1140 -DSNRTSTSNHNRIRQLRQEFQAKODEVEYRRRTTSEQOWPNARPAQYSGRHSVSE 1198
QY 1201 VOMOROROEERESSQOARQOYSSLPROSRKNASVSQDSWEQNTSPGEGFOGAKENPRYS 1260
DB 1199 VOYQOROEERESSQOARQOYSSLPROSRKNASVSQDSWEQNTSPGEGFOGAKENPRYS 1258
QY 1261 SYOGSRNGYLGSHGFNARVMELOELRQOEORRKEQOMKOPSEGPSNYDSYKRVODPS 1320
DB 1259 SYOGSRNGYLGSHGFNARVMELOELRQOEORRKEQOMKOPSEGPSNYDSYKRVODPS 1304
QY 1321 YAPKGPFRQOVPPSPSOVARLNRLQTEPKGRFPYS 1356
DB 1305 --YKGPFRQOVPPSPSOVARLNRLQTEPKGRFPYS 1337
RESULT 5
099NH2 PRELIMINARY: PRT: 1333 AA.
ID 099NH2:
AC 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PAR-3 180 KDA ISOFORM.
GN PAR3 OR PAR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH SWISS;
RX MEDLINE=99121117; PubMed=9920925;
RA Lin D., Gish G.D., Songyang Z., Pawson T.;
RT "The carboxyl terminus of B class ephrins constitutes a PDZ domain
blinding motif.";
RL J. Biol. Chem. 274:3726-3733(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH SWISS;
RX MEDLINE=20394297; PubMed=10934475;
RA Lin D., Edwards A.S., Fawcett J.P., Mhamali G., Scott J.D., Pawson T.;
RT "A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and apKC
signaling and cell polarity.";
RL Nat. Cell Biol. 2:540-547(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH SWISS;
RA Lin D.C. Jr., Pawson T.J.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY026057; AAK07669.1; -
DR HSSP; Q12923; 3PDZ.
DR MGD; MGI:2135608; Pard3.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 1333 AA; 149060 MW; AF67825C66DCE86 CRC64;

Query Match 89.7%; Score 6310.5; DB 11; Length 1333;
Best Local Similarity 89.7%; Pred. No. 0;
Matches 1216; Conservative 60; Mismatches 57; Indels 23; Gaps 4;

QY 1 MKVTVCFGFRTRVVPCGCGHMKVFSLIQOAVTRYRKAKAKDPNMYIQVRLHSGDGILD 60
DB 1 MKVTVCFGFRTRVVPCGCGHMKVFSLIQOAVTRYRKAKAKDPNMYIQVRLHSGDGILD 60

QY 61 LDDILCDVADBDKRLVAVEDEQDPHHGCDGTSASTGTQSPETFGSELGNTNNVSAFQPYQ 120
DB 61 LDDILCDVADBDKRLVAVEDEQDPHHGCDGTSASTGTQSPETFGSELGNTNNVSAFQPYQ 120
QY 121 ATSEIEVTSPVLRAANPLHVRSSDPALIGLSTSVSDSNFSEEPKRNKPTRWSTTAGEL 180
DB 121 ATSEIEVTSPVLRAANPLHVRSSDPALIGLSTSVSDSNFSEEPKRNKPTRWSTTAGEL 180
QY 181 KONTAGSPPTCDKRNKENTRSLPRDTSNMSNOPORNASSLSASHPMKWLEKEQOE 240
DB 181 KONTAGSPPTCDKRNKENTRSLPRDTSNMSNOPORNASSLSASHPMKWLEKEQOE 240
QY 241 DGEEDNSRVEEYGHADTLEHINPNSLDDMYLVEVPNDGEGFLGTHVVPFASRGRTG 300
DB 241 EGEEDNSRVEEYGHADTLEHINPNSLDDMYLVEVPNDGEGFLGTHVVPFASRGRTG 300
QY 301 LLYKLEKGGKAEHNLFPBNDCTVIRINDGDLNRRFEQAOHMFROAMKTPITWVHPA 360
DB 301 LLYKLEKGGKAEHNLFPBNDCTVIRINDGDLNRRFEQAOHMFROAMKTPITWVHPA 360
QY 361 ANKEOYEOLSOSEKKNYYSRFSPOXYIDNRSVNSAGLHTVORAPRLNHPPOIDSHS 420
DB 361 ANKEOYEOLSOSEKKNYYSRFSPOXYIDNRSVNSAGLHTVORAPRLNHPPOIDSHS 420
QY 421 LPHSAHPSGKPPSPAPASAPQONFSTVSSQYNTKKIGKRLNLOLKKGTGEGSITSRDY 480
DB 421 LPHSAHPSGKPPSPAPASAPQONFSTVSSQYNTKKIGKRLNLOLKKGTGEGSITSRDY 480
QY 481 TTGGSAPITYVKNLIPRGAALIDGRKLACGDRLEVNGVDLVGKSQEEVYSLRSTKMGTY 540
DB 481 TTGGSAPITYVKNLIPRGAALIDGRKLACGDRLEVNGVDLVGKSQEEVYSLRSTKMGTY 540
QY 541 SLVVFQEDAFPHRELNAPSPOMQIPKRTKAEDEDVLVLPDGRREFLTFEVLNDGSGAG 600
DB 541 SLVVFQEDAFPHRELNAPSPOMQIPKRTKAEDEDVLVLPDGRREFLTFEVLNDGSGAG 600
QY 601 LGVSVKGNRSKENHADLGIIFYKSIINGGAASKDRRLRVNDQLIANGESILGKTNDAME 660
DB 601 LGVSVKGNRSKENHADLGIIFYKSIINGGAASKDRRLRVNDQLIANGESILGKTNDAME 660
QY 661 TLRRSSTEGCKRMGIQIYARISCKNELKSGSPGPELPLETALDDRRRISHSLYS 720
DB 661 TLRRSSTEGCKRMGIQIYARISCKNELKSGSPGPELPLETALDDRRRISHSLYS 720
QY 721 GIEGLDESPRNAALSIRINGESGYOLSPVNNPQDVTIETEDRLPYLPHLSDSSSS 780
DB 721 GIEGLDESPRNAALSIRINGESGYOLSPVNNPQDVTIETEDRLPYLPHLSDSSSS 780
QY 781 SHDVGFTVADAGTWAKAALISDADCSLSPDVYPVLAFOREGFGROSMSEKRTKQFSDAS 840
DB 781 SHDVGFTVADAGTWAKAALISDADCSLSPDVYPVLAFOREGFGROSMSEKRTKQFSDAS 840
QY 841 QLDFVYTRKSKSMDLGIADETKLTNYDDQAGSPSDVPGSLGKSSLSLSLQTVAAEY 900
DB 841 QLDFVYTRKSKSMDLGIADETKLTNYDDQAGSPSDVPGSLGKSSLSLSLQTVAAEY 900
QY 901 TLNGDIPFHRPRRIIRGRGCSNFPRAIDKSYDKRAVDDDDGCMETLEEDTESSRSGR 960
DB 901 TLNGDIPFHRPRRIIRGRGCSNFPRAIDKSYDKRAVDDDDGCMETLEEDTESSRSGR 960
QY 961 ESYVSTASDQPSHSLEROMNGNOEKDGTDRKKDKTEKKKDKKDKKAKKAKMLKGL 1020
DB 961 ESYVSTASDQPSHSLEROMNGNOEKDGTDRKKDKTEKKKDKKDKKAKKAKMLKGL 1020
QY 1017 GMEFRFGKHRRKDDKMEKMRKIQDSFTSEEDVRYMKEQERIOAKTREFRERQARERDY 1076
DB 1017 GMEFRFGKHRRKDDKMEKMRKIQDSFTSEEDVRYMKEQERIOAKTREFRERQARERDY 1076
QY 1081 AEIODEFHRTFGCDDDELMATGVSSYEGSMALNARPOS PREGHMMDALYAQVKRRNSKPS 1140
DB 1077 AEIODEFHRTFGCDDDELMATGVSSYEGSMALNARPOS PREGHMMDALYAQVKRRNSKPS 1135

```

QY 1141 VDSNFTSPSNHDIQRLROEFOAKODEVEDHRRRTYSFQWPNAPATOGSRHSVYE 1200
Db 1136 -DSNFTSPSNHDIQRLROEFOAKODEVEDHRRRTYSFQWPNAPATOGSRHSVYE 1194
QY 1201 VOMOROEERESQOQROYSSLPROSRRNASSVSQDSWBNYSGEGFOJAKNPRYS 1260
Db 1195 VOVOROEERESQOQROYSSLPROSRRNASSVSQDSWBNYSGEGFOJAKNPRYS 1254
QY 1261 SYGSRNGYIGGHGFNARVWLETOFELROEORKEEOMKOPPSGPNYSYKKVODPS 1320
Db 1255 SYGSRNGYIGGHGFNARVWLETOFELROEORKEEOMKOPPSGPNYSYKKVODPS 1300
QY 1321 YAPKGFRODVPSPSOVARINRLOTPEKGRFPYS 1356
Db 1301 ---VRGPFRODVPSPSOVARINRLOTPEKGRFPYS 1333

```

```

RESULT 6
Q9BY57 PRELIMINARY; PRT: 1031 AA.
ID Q9BY57;
AC Q9BY57;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN SHORT VARIANT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN 11;
RP SEQUENCE FROM N.A.
RA Fang C., Xu Y.;
RT "Exon/Intron Structure and Splicing Variants of a Novel Human Polarity
RT Gene, hASIP.":
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF196186; AAK27892.1; -
DR HSP; Q12923; 3pDZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SO SEQUENCE 1031 AA; 113418 MW; 40DD593636EAB999 CRC64;

```

```

Query Match 74.9%; Score 5268.5; DB 4; Length 1031;
Best Local Similarity 98.8%; Pred. No. 2e-290;
Matches 1022; Conservative 3; Mismatches 6; Indels 3; Gaps 1;

```

```

QY 1 MWTVCFGRTYVPCGCHMKVFSLLQOAVTRYKALAKDPNWIQVHLEHGDCGLD 60
Db 1 MWTVCFGRTYVPCGCHMKVFSLLQOAVTRYKALAKDPNWIQVHLEHGDCGLD 60
QY 61 LDDILCDVADDKRLVAVFEDDPHGGDGTASSTQSPETFGSELGTNNYSAFQPY 120
Db 61 LDDILCDVADDKRLVAVFEDDPHGGDGTASSTQSPETFGSELGTNNYSAFQPY 120
QY 121 ATSELEWTPSVLRANMPLHVRSSDPALIGLSTVSQSNSSSEPRKPTWSTTAGFL 180
Db 121 ATSELEWTPSVLRANMPLHVRSSDPALIGLSTVSQSNSSSEPRKPTWSTTAGFL 180
QY 121 ATSELEWTPSVLRANMPLHVRSSDPALIGLSTVSQSNSSSEPRKPTWSTTAGFL 180
Db 121 ATSELEWTPSVLRANMPLHVRSSDPALIGLSTVSQSNSSSEPRKPTWSTTAGFL 180
QY 181 KONTAGSPKTCRRKKDENTYSLPRDTSNMSNOFORNANASSLSASHPMVKWLEKOEDE 240
Db 181 KONTAGSPKTCRRKKDENTYSLPRDTSNMSNOFORNANASSLSASHPMVKWLEKOEDE 240
QY 181 KONTAGSPKTCRRKKDENTYSLPRDTSNMSNOFORNANASSLSASHPMVKWLEKOEDE 240
Db 181 KONTAGSPKTCRRKKDENTYSLPRDTSNMSNOFORNANASSLSASHPMVKWLEKOEDE 240
QY 241 DGTFEDNSRVEPGHADDTGLEHINPNSLSDMYKLVFVPDGGPLGIHVPPSARGRTLG 300
Db 241 DGTFEDNSRVEPGHADDTGLEHINPNSLSDMYKLVFVPDGGPLGIHVPPSARGRTLG 300
QY 301 LTVKRLKGGKAHEHLEFRENOCIVINDGDLRNRRFOAOHMFQOARTPIIMFHYVPA 360
Db 301 LTVKRLKGGKAHEHLEFRENOCIVINDGDLRNRRFOAOHMFQOARTPIIMFHYVPA 360
QY 361 ANKEQYEOULSQSEKNNYSSRFPDSQYIDNRSVNSAGLHTVQAPRLNHPEDIDSHSR 420

```

```

Db 361 ANKEQYEOULSQSEKNNYSSRFPDSQYIDNRSVNSAGLHTVQAPRLNHPEDIDSHSR 420
QY 421 LPHSAHPGPPAPAPASAPQNFSTTVSSGYTKTKIGRLNIQLKKGEGIGFTSRDY 480
Db 421 LPHSAHPGPPAPAPASAPQNFSTTVSSGYTKTKIGRLNIQLKKGEGIGFTSRDY 480
QY 481 TIGGSAPYVKNILPRGAALIDGRKAGDRILEVNGVDLYKSGEEVYSLRSTKMGTV 540
Db 481 TIGGSAPYVKNILPRGAALIDGRKAGDRILEVNGVDLYKSGEEVYSLRSTKMGTV 540
QY 541 SLVFEROEDAFPRRLNEPSOMQIPKTKAEDDYLVTDPGTRFELTEFVPLNDGSGAG 600
Db 541 SLVFEROEDAFPRRLNEPSOMQIPKTKAEDDYLVTDPGTRFELTEFVPLNDGSGAG 600
QY 601 LGSVYKGRSKENHADLGIIFYKSTIINGGAASKDGRRLRVNDQLIANGESILKKTNDAME 660
Db 601 LGSVYKGRSKENHADLGIIFYKSTIINGGAASKDGRRLRVNDQLIANGESILKKTNDAME 660
QY 661 TLRRSMSTEGNRKGMIOIIVARRISKCNELKSPGSPGPELPIETALDRERRISHLYS 720
Db 661 TLRRSMSTEGNRKGMIOIIVARRISKCNELKSPGSPGPELPIETALDRERRISHLYS 720
QY 721 GIEGLDPSPRNAALSRIMGESGYOLSPVYVNPPODDYITIEDDLPLVPLPHLSQSSSS 780
Db 721 GIEGLDPSPRNAALSRIMGESGYOLSPVYVNPPODDYITIEDDLPLVPLPHLSQSSSS 777
QY 781 SHDVGFEVTPADCTWAKAIDSADCSLSPDVPVIAFORREGFGROSMEKRTKOFDAS 840
Db 781 SHDVGFEVTPADCTWAKAIDSADCSLSPDVPVIAFORREGFGROSMEKRTKOFDAS 837
QY 841 QLDPEVTRKSKMDIGIADETKLTAVDQKAGSPRDVPSLGITKSSLSLQTAVAEY 900
Db 841 QLDPEVTRKSKMDIGIADETKLTAVDQKAGSPRDVPSLGITKSSLSLQTAVAEY 897
QY 901 TLNGDIPFHRPRRIIRGCGNESPRALIDSYKPAVDDDEGEMTLEEDTESSSGR 960
Db 901 TLNGDIPFHRPRRIIRGCGNESPRALIDSYKPAVDDDEGEMTLEEDTESSSGR 957
QY 961 ESVSTASDQPSHSLEKONGQEKDTRKKDKTGKTKKDDKEDKKAKKMKLKL 1020
Db 961 ESVSTASDQPSHSLEKONGQEKDTRKKDKTGKTKKDDKEDKKAKKMKLKL 1017
QY 1021 GDMFRFGHRRDK 1034
Db 1018 GDMFSLAKLKEKR 1031

```

```

RESULT 7
Q96RM7 PRELIMINARY; PRT: 988 AA.
ID Q96RM7;
AC Q96RM7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ATYPICAL PKC ISOTYPE-SPECIFIC INTERACTING PROTEIN SHORT VARIANT
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN 11;
RP SEQUENCE FROM N.A.
RA Fang C., Xu Y.;
RT "Exon/Intron Structure And Splicing Variants of A Novel Human Polarity
RT Gene, hASIP.":
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF332592; AAK69192.1; -
SO SEQUENCE 988 AA; 108545 MW; 89F2139B06F7F7F CRC64;

```

```

Query Match 71.3%; Score 5014; DB 4; Length 988;
Best Local Similarity 94.7%; Pred. No. 5.2e-276;

```

Matches 979: Conservative 3; Mismatches 6; Indels 46; Gaps 3;				
Qy	1	MKVTVCEGRTRVYVPCGDMHKVFSLIQAAVTRYRKALAKDPNMYIQVHLEHGDCGLD	60	
Db	1	MKVTVCEGRTRVYVPCGDMHKVFSLIQAAVTRYRKALAKDPNMYIQVHLEHGDCGLD	60	
Qy	61	LDLILCDVADKDLVAVFDEODPHGGDGTASSTGTQSPEIFGSELTNNVSAFQYQ	120	
Db	61	LDLILCDVADKDLVAVFDEODPHGGDGTASSTGTQSPEIFGSELTNNVSAFQYQ	120	
Qy	121	ATSELEVPVYLRAVMPVHVRSSDPALIGSTSVSDSNFSESESRKNPTWSTAGFL	180	
Db	121	ATSELEVPVYLRAVMPVHVRSSDPALIGSTSVSDSNFSESESRKNPTWSTAGFL	180	
Qy	181	KONTAGSPKTCDDKKDEYRSLPRDTSNMSNOFORDNARSSLASHPVGVKLEKOEODE	240	
Db	181	KONTAGSPKTCDDKKDEYRSLPRDTSNMSNOFORDNARSSLASHPVGVKLEKOEODE	240	
Qy	241	DGTEEDNSRVEPVGHADTGLEHTPNFSLDDMYKLVLPNDGGPLGIVHVPESARGGRTLG	300	
Db	241	DGTEEDNSRVEPVGHADTGLEHTPNFSLDDMYKLVLPNDGGPLGIVHVPESARGGRTLG	300	
Qy	301	LLVRLKLEKGAHEHNLRENDCTVRINDGDLRNRFEQAQHMFRQAMKPTIWFHVVA	360	
Db	301	LLVRLKLEKGAHEHNLRENDCTVRINDGDLRNRFEQAQHMFRQAMKPTIWFHVVA	360	
Qy	361	ANKEQYEOLOSSEKNNYSRFPSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSR	420	
Db	361	ANKEQYEOLOSSEKNNYSRFPSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSR	420	
Qy	421	LPHSAHPGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLEKGTGEGSITSRDY	480	
Db	421	LPHSAHPGKPPSPAPASAPONVSTTVSSGYNTKKIGKRLNIOLEKGTGEGSITSRDY	480	
Qy	481	TTGSGAPLYVNNILPRGAIDODGLKAGDRLIEVNGVDLVGKSOEYVSLRSTMEGV	540	
Db	481	TTGSGAPLYVNNILPRGAIDODGLKAGDRLIEVNGVDLVGKSOEYVSLRSTMEGV	540	
Qy	541	SILVFERODAFHPRLENAEPESOMQIPKETEKADEEDIVLTPOGTREPLTFEYVPLNDGSGAG	600	
Db	541	SILVFERODAFHPRLENAEPESOMQIPKETEKADEEDIVLTPOGTREPLTFEYVPLNDGSGAG	600	
Qy	601	LGVSYKGRSKENHADLGIYFKYSIINGGAASFOGRLRYNDQLLIANGESLIGKTQDAME	660	
Db	601	LGVSYKGRSKENHADLGIYFKYSIINGGAASFOGRLRYNDQLLIANGESLIGKTQDAME	660	
Qy	661	TLRSMSTEGKRGMTIOLIVARISKCNELKSPGSPGPPELPIETALDDREKRISHSLYS	720	
Db	661	TLRSMSTEGKRGMTIOLIVARISKCNELKSPGSPGPPELPIETALDDREKRISHSLYS	720	
Qy	721	GIEGLDESPSRNAALSRLMGESGYOLSPVYVMPQDDTVIIEDDRPLVLPRLSDQSSSS	780	
Db	721	GIEGLDESPSRNAALSRLMGESGYOLSPVYVMPQDDTVIIEDDRPLVLPRLSDQSSSS	780	
Qy	781	SHDVGFTADAGTAAKAAISDSADCSLSPVDVYLAFORBEGFGQSMSEKTKOPSDAS	840	
Db	781	SHDVGFTADAGTAAKAAISDSADCSLSPVDVYLAFORBEGFGQSMSEKTKOPSDAS	840	
Qy	841	QLDFKTKRKSMDGLIETKLTNTVDDOKAGSPRDVGPGLTKSSLSLESLQTAFAVEV	900	
Db	841	QLDFKTKRKSMDGLIETKLTNTVDDOKAGSPRDVGPGLTKSSLSLESLQTAFAVEV	900	
Qy	901	TLNGDIPHRPRRLIRRGCSNESFRRAIDKSYDKPAVDDDEGMETLEEDTEESRSRGR	960	
Db	901	TLNGDIPHRPRRLIRRGCSNESFRRAIDKSYDKPAVDDDEGMETLEEDTEESRSRGR	960	
Qy	961	ESVSTAASQPSHSLEROMANGOEKDKTDKDKDTGKEKKKDKDKKKKKMKLKL	1020	
Db	961	ESVSTAASQPSHSLEROMANGOEKDKTDKDKDTGKEKKKDKDKKKKKMKLKL	1020	
Qy	1021	GDMFFGKHKRDKD 1034		
Db	1021	GDMFFGKHKRDKD 1034		
Qy	975	GDMFSLAKLKEPKR 988		
Db	975	GDMFSLAKLKEPKR 988		

RESULT	8	
Q96K28	PRELIMINARY:	PRF: 865 AA.
AC	Q96K28:	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	CNDA FLJ14829 FIS, CLONE OVARC1000945, MODERATELY SIMILAR TO	
DE	RATTUS NORVEGICUS ATYPICAL PKC SPECIFIC BINDING PROTEIN.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N. A.	
RC	TISSUE=OVARIAN CARCINOMA;	
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,	
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,	
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,	
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,	
RA	Watanabe S., Kimura K., Murekami K., Ishi S., Kawai Y., Saito K.,	
RA	Yanamoto J., Makamatsu A., Nakamura Y., Nagahari K., Masuno Y.,	
RT	Niomiya K., Iwayanagi T.;	
RL	"NEO human cDNA sequencing project."	
DR	Submitted (May-2001) to the EMBL/Genbank/DBJ databases.	
DR	EMBL: AK027735; BAB5330.1; -	
SQ	SEQUENCE 865 AA: 95161 MW; 5378BBCD406D0835 CRC64;	
Query Match 62.0%; Score 4361; DB 4; Length 865;		
Best Local Similarity 94.8%; Pred. No. 4,5e-239;		
Matches 852; Conservative 5; Mismatches 8; Indels 34; Gaps 2;		
Qy	136	MLVHRRSSDPALLIGLSTVSQSDSNFSSEPSRKNPTWSTAGFLKONTAGSPKTCDDRK
Db	1	MLVHRRSSDPALLIGLSTVSQSDSNFSSEPSRKNPTWSTAGFLKONTAGSPKTCDDRK
Qy	196	DENYRSLPDDTSNMSNOFORDNARSSLASHPVGVKWLKODEGTEEDNSRVEPVGH
Db	61	DENYRSLPDDTSNMSNOFORDNARSSLASHPVGVKWLKODEGTEEDNSRVEPVGH
Qy	256	ADTGLEHTPNFSLDDMYKLVLPNDGGPLGIVHVPESARGGRTLGILVRLKLEKGAHEH
Db	121	ADTGLEHTPNFSLDDMYKLVLPNDGGPLGIVHVPESARGGRTLGILVRLKLEKGAHEH
Qy	316	NLFRENDCTVRINDGDLRNRFEQAQHMFRQAMKPTIWFHVPAANKQYEOLOSSEKN
Db	181	NLFRENDCTVRINDGDLRNRFEQAQHMFRQAMKPTIWFHVPAANKQYEOLOSSEKN
Qy	376	NYSSRFPSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSRILPHSAHPGKPPSP
Db	241	NYSSRFPSPSOYIDNRSVNSAGLHTVQAPRLNHPPEQIDSHSRILPHSAHPGKPPSP
Qy	436	ASAPONVSTTVSSGYNTKKIGKRLNIOLEKGTGEGSITSRDYITGSGAPLYVNNILP
Db	301	ASAPONVSTTVSSGYNTKKIGKRLNIOLEKGTGEGSITSRDYITGSGAPLYVNNILP
Qy	496	RGAAIODGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGSTVSLVFRQODAFHPR
Db	361	RGAAIODGRLKAGDRLIEVNGVDLVGKSOEYVSLRSTKMGSTVSLVFRQODAFHPR
Qy	556	LNAEPSONOIPKETADEEDIVLTPOGTREPLTFEYVPLNDGSGAGVGVKGRSKENNA
Db	421	LNAEPSONOIPKETADEEDIVLTPOGTREPLTFEYVPLNDGSGAGVGVKGRSKENNA
Qy	616	DGIFVYSIINGGAASKDRLRYNDQLLIANGESLIGKTQDAMEETLRRSMSTEGKRG
Db	481	DGIFVYSIINGGAASKDRLRYNDQLLIANGESLIGKTQDAMEETLRRSMSTEGKRG
Qy	676	IOLIVARRISKCNELKSPSPGPPELPIETALDDREKRISHSLYSGLDELDSPSRNAL
Db	735	IOLIVARRISKCNELKSPSPGPPELPIETALDDREKRISHSLYSGLDELDSPSRNAL

```

Db 541 IOLIVARRISKCNELKSPGSPGPPELPIETALDRERRISHSLYSGIGLDESPRNAL 600
QY 736 SRIMGESGKYQLSPYVNMPODDVTIIEDDRILVLPPLHSDOSSSSSHDDVGFYADAGTW 795
Db 601 SRIM---GKQYLSPTVNMPODDVTIIEDDRILVLPPLHSDOSSSSSHDDVGFYADAGTW 657
QY 796 AKKAISDSADCSLSPVDVYLAFOREGFGROSMSEKRTQFSDASQDLFVTKRKSMDL 855
Db 658 AKKAISDSADCSLSPVDVYLAFOREGFGRO----- 689
QY 856 GIADETKLTNTVDOKAGSPSRDVGPSLGLKSSSLESLQTAVAEYTLNGDIPFHRPRRI 915
Db 690 ---DETKLTNTVDOKAGSPSRDVGPSLGLKSSSLESLQTAVAEYTLNGDIPFHRPRRI 746
QY 916 IGRGNCNESFRAIDKSTDKPAVDDDEGMETLEDETESSRSRGESVSTASQPSHLE 975
Db 747 IGRGNCNESFRAIDKSTDKPAVDDDEGMETLEDETESSRSRGESVSTASQPSHLE 806
QY 976 ROMNGNOERKGDTRKRRKKTGKKKDRKDKKAKKMGKGLGDMFRFGKRRDK 1034
Db 807 ROMNGNOERKGDTRKRRKKTGKKKDRKDKKAKKMGKGLGDMFSLAKLPEKR 865

```

```

RESULT 9
09HC48 PRELIMINARY; PRT; 667 AA.
AC 09HC48;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CCLL TUMOR ANTIGEN SE2-5 (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=21143360; PubMed=11149944;
RA Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,
RA Schendendorf D.;
RT "Serological detection of cutaneous T-cell lymphoma-associated
RT antigens.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
DR EMBL; AF177228; AAG33676.1; -.
DR HSSP; 012923; 3PZ.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
FT NON_TER 1
FT TER 667
SQ SEQUENCE 667 AA; 73499 MW; C653EC16802BAE02 CRC64;

```

```

Query Match 47.4%; Score 3335; DB 4; Length 667;
Best Local Similarity 94.3%; Pred. No. 4,7e-181;
Matches 657; Conservative 3; Mismatches 7; Indels 30; Gaps 1;

```

```

QY 313 EHEULFPRENCIVRINDGDLRNRFRFOAOMFRQAMRPTLIFHVVPANKEQYBOLSSQ 372
Db 1 EHEULFPRENCIVRINDGDLRNRFRFOAOMFRQAMRPTLIFHVVPANKEQYBOLSSQ 60
QY 373 EKNYNTSSRSPDSQYIDNRSVSAGLHTVORAPRLNHPPEQIDSHSLPHSAHPSGKRP 432
Db 61 EKNYNTSSRSPDSQYIDNRSVSAGLHTVORAPRLNHPPEQIDSHSLPHSAHPSGKRP 120
QY 433 SAPASAPONFSTVSSGNTKKIGKRLNIQLKGTGEGITSRDVYIGSAPITYKN 492
Db 121 SAPASAPONFSTVSSGNTKKIGKRLNIQLKGTGEGITSRDVYIGSAPITYKN 180
QY 493 ILPRGAIDGRLKAGDRIEYNGVDLGKSGEEVYSLRSTKKEGVYSLVFRQEDAFH 552
Db 118 ILPRGAIDGRLKAGDRIEYNGVDLGKSGEEVYSLRSTKKEGVYSLVFRQEDAFH 143

```

```

Db 181 ILPRGAIDGRLKAGDRIEYNGVDLGKSGEEVYSLRSTKKEGVYSLVFRQEDAFH 240
QY 553 PRELNAEPQOMQIPRETKADEDEDIVLTPDGTREFTFEPVLPNDGSGAGLGVYAGNSKE 612
Db 241 PRELNAEPQOMQIPRETKADEDEDIVLTPDGTREFTFEPVLPNDGSGAGLGVYAGNSKE 300
QY 613 NHADGIFVKSILINGGAASKDGRRLVNDQLAVNGESLGTNDODAMETLRRSKTEGKN 672
Db 301 NHADGIFVKSILINGGAASKDGRRLVNDQLAVNGESLGTNDODAMETLRRSKTEGKN 360
QY 673 RCMOLIYARRISKCNELKSPGSPGPPELPIETALDRERRISHSLYSGIEGLDESPRN 732
Db 361 RCMOLIYARRISKCNELKSPGSPGPPELPIETALDRERRISHSLYSGIEGLDESPRN 420
QY 733 AALSRIMGESKYQLSPYVNMPODDVTIIEDDRILVLPPLHSDOSSSSSHDDVGFYADAGTW 792
Db 421 AALSRIMGESKYQLSPYVNMPODDVTIIEDDRILVLPPLHSDOSSSSSHDDVGFYADAGTW 480
QY 793 GTWAKKAISDSADCSLSPVDVYLAFOREGFGROSMSEKRTQFSDASQDLFVTKRKS 852
Db 481 GTWAKKAISDSADCSLSPVDVYLAFOREGFGRO----- 514
QY 853 MDLGIADETKLTNTVDOKAGSPSRDVGPSLGLKSSSLESLQTAVAEYTLNGDIPFHRPR 912
Db 515 ---IADETKLTNTVDOKAGSPSRDVGPSLGLKSSSLESLQTAVAEYTLNGDIPFHRPR 570
QY 913 PRIIRGNCNESFRAIDKSTDKPAVDDDEGMETLEDETESSRSRGESVSTASQPSH 972
Db 571 PRIIRGNCNESFRAIDKSTDKPAVDDDEGMETLEDETESSRSRGESVSTASQPSH 630
QY 973 SLEROMNGNOERKGDTRKRRKKTGKKKDRKDKK 1009
Db 631 SLEROMNGNOERKGDTRKRRKKTGKKKDRKDKK 667

```

```

RESULT 10
096NX7 PRELIMINARY; PRT; 1205 AA.
AC 096NX7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICED VARIANT P.
GN PAR3L.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gao L., Macara I.G., Joberty G.;
RT "Multiple splice variants of Par3 and of a novel related gene, Par3L,
RT produce functionally different proteins.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF426250; AAL30664.1; -.
SQ SEQUENCE 1205 AA; 132494 MW; 26E6704CCDCE8CD8 CRC64;

```

```

Query Match 29.3%; Score 2061.5; DB 4; Length 1205;
Best Local Similarity 37.9%; Pred. No. 1.7e-108;
Matches 539; Conservative 191; Mismatches 377; Indels 317; Gaps 49;
QY 1 MKVTYFGRTGVVPCGSHMKVFSLIDAVYRKAIAKPNWYIOVHREHDDGSLD 60
Db 1 MKVTYFGRTGVVPCGSHMKVFSLIDAVYRKAIAKPNWYIOVHREHDDGSLD 60
QY 61 LDDIICDVADDDKRIYVAYDDPDDPHHGGDGSASSTGTQSPFIRGSELGNVSAFQPYQ 120
Db 61 PDDVADVDEDKILAYFEDEDEPLHKIESPSGNPADROSADAFETEVAA-QLAAFRP-- 117
QY 121 AATSELEVPSTVLRANMPLHVRSSDPALIGLSTSVSDNSFSSEPSKKNPTKSTAGFL 180
Db 118 IGELEIVTPSALKIGTPIVLRSSDP----- 143

```

```
QY 181 KONTAGSPKTCRDKDENRSLPRTSNMNSNOFORNARSSLSASHPMVKMLEQCODE 240
Db 144 ---VGP-----ADTOPSASHP-CGOSLKLVPDS 170
QY 241 DGTEDNSRVEPVGHADTGLEHLPNF--SLDDMVKLYEVNDPNDGGLGIHVVP-FSARGR 297
Db 171 TONLEDR--EVLNGVQTELLTSPRTKDTLSDMTVEISGGGGLGIHVVPFSSLSGR 227
QY 298 TGLGLVRLKLEKGAHEHNLFRENDCIVRINDGLNRRFEOAQHMFQAMRTPIIMFHV 357
Db 228 ILGLFRIGIEDNSRSKRGLFHEHNCIYKINNVLDVDTFAOAOVFOAKMSPVLLHV 287
QY 358 VPANKEOYEOLSOSEKNYSSRSPDSQYIDNRSVNSAGLHTVORARLHNPPEQIDS 417
Db 288 LPPONREQYER-SVIGSLNIFGNNDGVLTQVPPVHGKSGLKT--ANLTGTDSPENDA 343
QY 418 HSRLPASHAPS-----GKPPSAPASAPQNVFTYSSGYNTKKIGRLNIQLKKGTEGLGF 473
Db 344 SASLQONKSPRVPRLGKSPSSLSLSP-----LMGFSKNNAKIKIDLKKGPEGGLGF 395
QY 474 SITSRDVTIGSAPITYVNIILPRGAIIODGRLKAGDRLIEVNGVDLVGKSQEBVSVLSR 533
Db 396 TVVTRDSSIHGBPIFVANNILPRKGAIKDGRLOSGDRILEVNGRVDYGTQDELYAMLSR 455
QY 534 TRMESTVLLVPRODAFHPRELNAPSOMQIPEKTKAEDIEDIYLTPDGTREFTLEVPVL 593
Db 456 TKOGETASLVIAROGHFLPRELKEPDCALSTETS-----EQLTFEIDL 501
QY 594 NDSGSGAGVSVKGNRSKEMNADGIFVKSTINGGASKDGRLVNDOLIANNGSLGK 653
Db 502 NDSGSGAGVSVKGNRSKEMNADGIFVKSTINGGASKDGRLVNDOLIANNGSLGK 561
QY 654 TNODAMETLRSMSTEGNKGRIQILVARISKCMELKSPGSPPELPIETALDRERR 713
Db 562 SNHEMETLRSMSTEGNKGRIQILVARISKCMELKSPGSPPELPIETALDRERR 596
QY 714 ISHSLSIGIEIDESPSSNALSRIKMGESGYQLSPYVNMPODQVITIEDRLVLPPLH 773
Db 597 -----EDPACGAFSKPCFEN--CONAVTTSRRDINSI-----L 628
QY 774 SDQSSSSHD-DVGFVPTADAGTMAKAISDS-----ADCSLSPDVDPVLAFORE 821
Db 639 HPLGCSQODKOKGILLPNDG-WAISEVPPSPPTPSALGLGLEDSHSGSAVAYFPQ 687
QY 822 GFGROSMEKRTKOPSDASOLDVFVTRKSKSMDLGIADETKLTNTVDDKAGSPSRDVGPS 881
Db 688 HINFRSVTPAR-----QPEISILKASKSMDL--VPDESKVHSLAGKSESPSKDFGPT 738
QY 882 LGLKSSSLESILQIVAAEVTLLNGDIPEHPRPRITRGCNESPRAAIDKSYDKPAVDD 941
Db 739 LGLKSSSLESILQIVAAEVTLLNGDIPEHPRPRITRGCNESPRAAIDKSYDKPAVDD 792
QY 942 DEGMTLEED--TESSSRGRESVSTASDQSHSLEROMNGQEGDGT---DRKKK-- 994
Db 793 -----EETADGISDSSSHSGGALNCESPOGNSLEDEMENKARVYKTKKEKKEKKEK 848
QY 995 -TGKSKKDRKEDKMKAKKMLKGLDMFRFGKHKRD--DKIEKTKIKIOESFTSEE 1051
Db 849 LKVKKKKKRENEDEPERIKK--KGFGAMLRFGKKKEDKGKAKOKGLK--HGLGRE 903
QY 1052 EIRIKKOEROIQAOTREFREKQARE-RDYA-----ETQDEHRTGCDDEL--MYGVSS 1103
Db 904 ELEKKKEERERIGAKHOETIREKQARGLLDYATGALIGSYVDMO-----DDEMPNAYRVNH 958
QY 1104 Y-EGSMALN-----ARPOS-----PREGH-----MMDALAYOKKPRNSKSPVDSNR 1145
Db 959 FREPCTSANVFSPPRPGRPGYPRDGHPLSPERDHLGLTAKVKNKPRHPL--VPADSGR 1017
QY 1146 SPPSNHDIRIOLRQEOQAKOD-----EDVEDRRRTYSFEQPMAPARPAQTQSGRHSVYE 1200
Db 1018 PFGGSDIRIQIKRKEYQAARRGFPFLYEDDEGRARPSEYDLDLWVGR-CPDONAHNLARE 1076
```

```
QY 1201 VOMORQOBERESSQOARQYSILPROSKKNASSVSQDSWEONYSFG--EGFQSAKENR 1258
Db 1077 -----GMEHOYASLPRGPAD-----PVDYLPAPRGLYKKERELPL 1112
QY 1259 YSSYGSRNGYLGGHFNARVWLETOELLRQORRKEQOMKQOPSEGSNNDYKKYOD 1318
Db 1113 YP-----GAHPM-----PPKGSYPRPTELRVAD 1136
QY 1319 ---PSYAPP-----KGFPRQDVPSPSQVARNLQTPREKGRP 1353
Db 1137 LRPQHYPPRRPAQHKGFPRQDVPSPSQVARNLQTPREKGRP 1178
RESULT 11
Q96NX6 PRELIMINARY; PRT; 1143 AA.
AC 096NX6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PARTITIONING-DEFECTIVE 3-LIKE PROTEIN SPLICER VARIANT B.
GN PARL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Gao L., Macara J.G., Joberty G.;
RT "Multiple splice variants of par3 and of a novel related gene, par3L,
RT produce functionally different proteins."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF428251; AAL30665.1;
SQ SEQUENCE 1143 AA; 126102 MW; 368B69CF81D45E7E CRC64;
Query Match 25.9%; Score 1825.5; DB 4; Length 1143;
Best Local Similarity 35.0%; Pred. No. 3.8e-95;
Matches 498; Conservative 184; Mismatches 363; Indels 379; Gaps 49;
QY 1 MKVTQCFGRFVRVPCGDMKVFSLIOQAVMYRKAIKDNPWYQVRLHGGGILD 60
Db 1 MKVTQCFGRFVRVPCGDMKVFSLIOQAVMYRKAIKDNPWYQVRLHGGGILD 60
QY 61 IDDIICDVADDDRLVAAYFDEODPHHGCGDGTASSTGTGSPETIFGSELGTNNVSAFQPYO 120
Db 61 PDVVLADVVEDDKLVAVEEBOEPLKIESPGNPRADRSPPAFIEVAA--QLAAKP-- 117
QY 121 ATSEIEVTVPILRANKPLHVRSSDPALIGLSTVSDSNFSSEEPSRKNPRTMSTTAGFL 180
Db 118 ICGEIEVTPSALKLGTPLLYRRSSDP----- 143
QY 181 KONTAGSPKTCRDKDENRSLPRTSNMNSNOFORNARSSLSASHPMVKMLEQCODE 240
Db 144 ---VGP-----ADTOPSASHP-CGOSLKLVPDS 170
QY 241 DGTEDNSRVEPVGHADTGLEHLPNF--SLDDMVKLYEVNDPNDGGLGIHVVP-FSARGR 297
Db 171 TONLEDR--EVLNGVQTELLTSPRTKDTLSDMTVEISGGGGLGIHVVPFSSLSGR 227
QY 298 TGLGLVRLKLEKGAHEHNLFRENDCIVRINDGLNRRFEOAQHMFQAMRTPIIMFHV 357
Db 228 ILGLFRIGIEDNSRSKRGLFHEHNCIYKINNVLDVDTFAOAOVFOAKMSPVLLHV 287
QY 358 VPANKEOYEOLSOSEKNYSSRSPDSQYIDNRSVNSAGLHTVORARLHNPPEQIDS 417
Db 288 LPPONREQYER-SVIGSLNIFGNNDGVLTQVPPVHGKSGLKT--ANLTGTDSPENDA 343
QY 418 HSRLPASHAPS-----GKPPSAPASAPQNVFTYSSGYNTKKIGRLNIQLKKGTEGLGF 473
Db 344 SASLQONKSPRVPRLGKSPSSLSLSP-----LMGFSKNNAKIKIDLKKGPEGGLGF 395
QY 474 SITSRDVTIGSAPITYVNIILPRGAIIODGRLKAGDRLIEVNGVDLVGKSQEBVSVLSR 533
```

```

Db      396 TVVTRUSSINGPPIFYVKNLIPKGAIAIKGRIQSDRIIEVGRVOTGTQDEIYAMLR 455
      534 TKMGCTVSLVFEKQDAFHFREINAEPSQMIPKTKADEIDIVLTPTGTRFLFEVPL 593
      456 TKQGETASLVIANQEGHFLREL----- 478
      594 NDGSGAGLVSVKGNRKNENHADLGIFVKSIIINGCAASKDGRILRVNDOLIAVNGSLI 653
      479 -----DGRIRMDQOLIAVNGSLI 499
      654 TTNODAMETLRSMSTEGNKGMIOLIVARRISKCNELKSPGPELPLEVALDRERR 713
      500 SNHEMFRLRSMWEGNIRGMIOVLIR-----PERRP----- 534
      714 ISHSLVSGIEGDESPSNALSRIMESGKYOLSPYVNNPODVTYIIEDRLPYLPHL 773
      535 -----EDPAECGAFSKPCPEN--CQNAVTTSRNDNSI-----L 566
      774 SDQSSSSHD-DVGFVTADACTAKAALSDS-----ACSLSPVDVPLAFORE 821
      567 HPLGTGSPQDKGILLIPNDG-WAESEVPSPPTPHSALGLCLEIYSHSGVDSAVYEPDQ 625
      822 GFGROSMSEKRTKOFSDASOLDVFKTRKSKSMDLGIADETKLNTVDOKACSPSDVGPS 881
      626 HINFRVTPAR-----QPSINLKASKSMDL-VPESEKSHSLAGOKSPSPKDEGPT 676
      882 LGLKSSSLESIGTAAVEYTLNGDIPFHRPRRIIRGCGNCFRPAIDKSYDPAYDD 941
      677 LGLKSSSLESIGTAAVEYRN-DLPFHRPRHAWGRCNCFRPAIDKSYDPAYDD 730
      942 DEGMETLEED--FEESRSGRESVSTASDQPSHLEKRONQOEGDKT--DRKKK-- 994
      731 -----EELFAGLSKSSHSGOGALNCEASAPQNSLEDENKARKYKKEKKEK 786
      995 -TGKSKKDRDKEDKMKKAKGMLKGLDMFRFGKHKRD--DKIEKTGKIKIOESFTSEE 1051
      787 LKVKKKKKKENEDPERIKK--KGFAMLRFGKKKEDGKAEQKTLK--HGLRRE 841
      1052 EIRIRKQOEIRIOAKTRFEFRQARE-RDYA-----EIODHRTFGCDEL--MGCVSS 1103
      842 ELEKKKEERIRGAKHOLIRKQARGLDYATGALISVYDMD-----DDMDNPAIRVNH 896
      1104 Y-EGSMALN-----ARPOS-----PREGH-----MMDALYAOVKKPRNSKSPYDSNR 1145
      897 FHEPCTSANVFRSPRPAPGPGYPRDGHPLSPERDILEGITAYVKNKPYHPL-VPADSGR 955
      1146 STPSNHDRIORLROEFOAKOD-----EDVEDRRRTYSFEOPWFNAPRATQSGHVSVE 1200
      956 PTGGSTDRIOKLKRYQARRRGFPPLYEDDEGRAPSEYDILWVGR-GPDGNANHLRFE 1014
      1201 VOMQROROEERESSQAOQOYSSLPQROSKNASSVSDSWONTSPG--EGFQOSAKENPR 1258
      1015 -----GMEHOYASLPRGPA-----PYDYLPAAPRGLYKEREPLY 1050
      1259 YSSVQSRNGYLGCHGFNARVMLETOELLROBORRKEQOMKOPRSEPSNYDSKKYVD 1318
      1051 YP-----GHNPRH-----PPKSYPRPTELKRVAD 1074
      1319 ---PSYAP-----KGFPODYPPSPSOVYARLNLQTPKGRP 1353
      1075 LRYOHYPRPAPRQHKGRFQODVPPSPPOHNPATQ--ETGRP 1116

```

```

RESULT 12
Q9NM14 PRELIMINARY; PRT; 347 AA.
AC 09NM14;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ20754 FIS, CLONE HEP02246 (UNKNOWN) (PROTEIN FOR
MGC:19518).

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Odayashi M.,
RA Nishi T., Shihahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RA "NEDO human cDNA sequencing project."
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RP TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK00761; BAA91366.1; --
DR EMBL; BC011711; AAH1711.1; --
DR EMBL; BC011711; AAH1711.1; --
SQ SEQUENCE 347 AA; 40538 MW; BE2B3557996EC91E CRC64;

```

```

Query Match 25.9%; Score 1822; DB 4; Length 347;
Best local similarity 100.0%; Pred. No. 1.2e-95;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      1010 MKAKKGMKLGDMFRGKHKHKKDKIEKTKIKIOESFTSEERIRMKOEROIKATRE 1069
      1 MKAKKGMKLGDMFRGKHKHKKDKIEKTKIKIOESFTSEERIRMKOEROIKATRE 60
      1070 FREORARENDYAEIODFHRFTGCDDELTYGCVSSYBSGSMALNARPOSPRGHMDALYNO 1129
      61 FREORARENDYAEIODFHRFTGCDDELTYGCVSSYBSGSMALNARPOSPRGHMDALYNO 120
      1130 VKKPRNSKSPYDSNRSTPSNHDRIORLROEFOAKODEVEDERRRTYSFEOPWPNARPA 1189
      121 VKKPRNSKSPYDSNRSTPSNHDRIORLROEFOAKODEVEDERRRTYSFEOPWPNARPA 180
      1190 TOSGRHSVSEVOMQROROEERESSQAOQOYSSLPQROSKNASSVSDSWONTSPGEG 1249
      181 TOSGRHSVSEVOMQROROEERESSQAOQOYSSLPQROSKNASSVSDSWONTSPGEG 240
      1250 FOSAKENPRYSYVQSGRNGLGCHGFNARVMLETOELLROBORRKEQOMKOPRSEGPSN 1309
      241 FOSAKENPRYSYVQSGRNGLGCHGFNARVMLETOELLROBORRKEQOMKOPRSEGPSN 300
      1310 YDSYKKTQDPSYAPKGFPODYPPSPSOVYARLNLQTPKGRPYS 1356
      301 YDSYKKTQDPSYAPKGFPODYPPSPSOVYARLNLQTPKGRPYS 347

```

```

RESULT 13
Q96DK9 PRELIMINARY; PRT; 624 AA.
AC 096DK9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FLJ25236 FIS, CLONE STM02096 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GASTRIC MUCOSA;
RA Ninomiya K., Wagatsuna M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RT Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.

```



```

Db 225 SSGNHOPFASGRISMOFLGDCNGYKMEAEAKLONQPAQOYQOQSHHAGHONGAYS 284
OY 246 -----DNRSEVYGHADTGLEHLPNFSLDMMKYLEVNDGGLGHHVPPSARGRTL 299
Db 285 SKSLPRSKREPELGQAYESIRE-----KDGEMLLINIEYSGDLTALPRKEHNG--- 335
OY 300 GILYKLEKGAKEHMLFRENDQIVRINDGDELRNRFEQAOHMFQAMFRTPIYHFVY 359
Db 336 GILYOHVEPGRRAEGR-LRDDRLILNIGIKILIGLESQVQOLRALE----- 384
OY 360 AANKQOYFOLQSEKNNYSSRSFSDQYIDNKNVSNAGLTHQVAPRLMHPPEQIDSHS 419
Db 385 -----SSELRYVRLKGRNR-----RQORDSKVAENEVAIYS----- 417
OY 420 RLPHSAHPSGKPPAPAPAPQNVFSTVSSGYNTKIKIGKRLNIOLKKGTEGLGFSITSRD 479
Db 418 -----PTRKHAAPVGTSLQV-----AMTRKIGRKTEIMLKKGPGNGLGFSVTRD 462
OY 480 VTTGSAPIYVKNILPRGAIDGRLKAGDRLIEVNGVDYLGKSOEEVVSILRSTKMEGT 539
Db 463 NPAGGHPIYIKNILPRCAIIEGRILKPDRLLEVDTGTPMGTQYDVVAILRGMAGAT 522
OY 540 VSLIVFRQ-----BDAPHPRELNA-----EPS----- 561
Db 523 VRTVYSQOQLAEQADQAPKSGAVAPVAPVAPAAAAPAPPYQKSSASRLFTH 582
OY 562 --QMOIYPR-----ETKAEDEDIVLPDG--PREFLTPVPLNDGSGAGIAGVYK- 607
Db 583 QOQOQLMESOHFIAGSESNASNDSLPSSNSWHSREELTLHPHDIEMKALGVSYK 642
OY 608 -----NRKENHADIGIPIYKSLINGGAASKDRLRVNDQILYANGESIL 651
Db 643 TCSNLMASGSSASSGSLKAKHODGLGIEYKANVHGGAAASRQGRILMNDQLSLVNGVSLR 702
OY 652 GKTNDAMETLRSM-STEGNKRMIOILIYARRLSK----- 687
Db 703 GQNNAEAMETLRAMVNTPGKHPTTLVGRKLKRASSDILLDSNSHSHSSNGG 762
OY 688 -----NELKSPGSPGELPTEALDRERRISHLSYSGIEGLDSSPNALSR 738
Db 763 SNSNSGNNNNSSNASNSGATYILSPERKROCGNGGGGAGNEMRMSNPVLDL 822
OY 739 MG-----ESKQYOLSPYVNMPO----- 755
Db 823 TGGICSSNSAOPSSQOOSHQOOPHSQOQOORLPAAPVCSAALRNEYMATNDWSP 882
OY 756 -----DDVILIEDRLVLP-----PLSDQSSSSHD-----DYGfV 788
Db 883 AQNHMLTAGHNTALLIEDAEPMSPTLPPARPIDGCHCTSSANPSQNLVANGOPPLNTY 942
OY 789 TADAGWAKAALSDADCSLSPDVPLAFORFGFQSGMSRK-----RTK 834
Db 943 PGTPTSSNDATAYSSQSLLETN-SCVEHFSDALGRSISIKHHAALDARETGYONK 1001
OY 835 QFSDASOLD-FYKTRKS-----KSMGLGLAD-----ETKLN-T 865
Db 1002 KLPERERERRIDLTSAVYGGISITARIASANAQFSGYKNAKTASITDEORQOLA 1061
OY 866 VDDQKASPRDVPSPGLKSSLSLSTAVAEVTLNGLDIFHRPRPILIRGCNBSF 925
Db 1062 AAEEARDOGLDGLPSLGMKSSLSLESLQTMVELOMS-DEPRHQALBARPGRGSESL 1120
OY 926 RAALDKSYD--K--AVDDDDGEMTELEETDESSRSRGSEVSASQPSHLSLERON 979
Db 1121 RAAVSEPDASKPRKTMLEDCD-----HEGFAASGRNG-----PROSSLN 1161
OY 980 GNOEGDTRDKKDTGKEKKDKDKKAKKMLKGLGDMFRGKHDKDKLEKTKG 1039
Db 1162 DGKH-GCKSSRAK-----KPSILRIGIHHFRGKARKNGVAVPVN 1200
OY 1040 ---KTIQOSFTSEERTIRMKQEDERIOAKTREFRERQOAREBYAELQDHRFGODEL 1096
Db 1201 YAVNISPTISVSTATSPQLQOQO-----QOLOHOQO----- 1235

```

```

OY 1097 MYGGVSYTEGSMALNARQSPREGHMDALYAOYKPRNSKSPVDSNRSTPSN----- 1150
Db 1236 -----QIPYAA--LAALERNGKPPYOPPPPLPAPNGVSGIHQND 1275
OY 1151 -----HRIQRLQOE--QOAKODEVEDERRRTYTFEOPMPARP 1188
Db 1276 IFNHRYOHTANTEDLHQOQHQOQHQHSHGRASARQOV-SMHSSTSGSGPSTAO 1334
OY 1189 ALOS-GRHSYS-----VEVOMORQOEFRESQOARQYSSLPROSRKNASSVSQDSE- 1241
Db 1335 QAQSNQVPRMSYEEYETVOOQRVGSIKSHSSA-----TSSSSPINVPHKA 1384
OY 1242 ---QNTSGEFGQSAKENPRYSYSGSRNGYLGHGFNARVLETOELLROE----- 1290
Db 1385 AAMNGYSPASINSAR-----SRGPFV-----TOYVITREGSGGIPAH 1422
OY 1291 --QRRKEQMKRQPS-----EGSNDYS 1312
Db 1423 LLQHQOQOOLQOQOPTYQYQKMSGPSQYGS 1453

```

```

RESULT 15
ID 096782 PRELIMINARY: PRT: 1464 AA.
AC 096782:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE BAZOOKA PROTEIN.
GN BAZ OR BAZOOKA OR CG5055.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuchinke U., Grave F., Knust E.;
RT "Control of spindle orientation in Drosophila by the Par-3-related
RT p2-domain protein Bazooka."
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ130871; CAI0224.1; -.
DR HSSP: Q12923; 3PDZ.
DR Flybase: FBgn000163; baz.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PS50106; PDZ; 3.
SQ
SEQUENCE 1464 AA; 157398 MW; 77381A5FCAAC4B CRC64;

```

```

Query Match 16.38; Score 1149.5; DB 5; Length 1464;
Best Local Similarity 26.28; Pred. No. 1.1e-56;
Matches 432; Conservative 212; Mismatches 470; Indels 537; Gaps 61;

OY 1 MKYTVCFGRTRVYVPCGGHMKVFSLIQAVTRYRKALADPNYVIOVHRLKGGDGLD 60
Db 1 MKYTVCFGRTRVYVPCGGHMKVFSLIQAVTRYRKALADPNYVIOVHRLKGGDGLD 58
OY 61 LDDILCDVADDDKRLVAVFDEOD-----PHNGGQTS-ASSTGSPSEIFGSELGTNVS 114
Db 59 PDCCVADVADDDKRLVAVFDEOD-----PHNGGQTS-ASSTGSPSEIFGSELGTNVS 116
OY 115 AFQPIYQ-ATSELEIVPVS-----VLRAVMPLVHRRSSDPALIGLTSVSDSNFSSEPSRK 169
Db 117 PTCPRDLSTPHIEVLTSTSGPAGAGLVGLMVRSSDPNLLA-SLKAEGSN----- 165
OY 170 PTFWSTTAGF-----LKONTAGSPKTCDDKNDN-----YRSLPRDTSNMS 210
Db 166 -KNSAARPHAGDSDPERFLFDLKAGGOLSPQWEEEDDSDSHOLKROLHQOQPHANGS 224
OY 211 NOFORDNARSSLASHMVG-----KMLEKQ--QDEGDETE----- 245

```

```

Db 225 SGNHQPARRSRLSMQFJGDGNGYKMMBAEKLQNOPPAQOTYYOOGSHHAGONGAYS 284
QY 246 -----DNSRVPVGHADTGLEHIFNPSLIDMVKLVEVPNDGPIGIIHVPPSARGRTL 299
Db 285 SKSLPRESKRKPELQAVYSIRE-----KQDEMILLINEYSPGLGTLPLPKERHGG----- 335
QY 300 GLLVRLKLEGGAEHBNLFRENDICVIRINDGLRNRREFQAOHMRQAMRTPIIMEHVVP 359
Db 336 GLLVHVEGSAERKQOL-RDDRILEINGIKLIGTESQVOEOLRALE----- 384
QY 360 AAMKFOYBOLSOSEKNYNYSSRFPDSOYIDNRSVNASGLHUYQARPLNHPPEQIDSMS 419
Db 385 ---SSELRYAVLIGDRNR-----ROORSKVADMVEATVS----- 417
QY 420 RLPHSHPSGKPPSPASAPQWVFSTVSSGYNTKIKGRRLNIOUKKEGLEFSTSD 479
Db 418 -----PFRKPHAAPVGSLOV-----ANTRLKIGKIQIMLAKGNGIGFSTYTHD 462
QY 480 VTIGSAPLYYKNILPRGAIODGLKACDRLILEVNGVDLVKSGQREVVSLRSTMEGT 539
Db 463 NPAGAHCPYIKNIIPRGALIEDGLKPGDRILEVDGTPMTGTQTDVVAIIRGMPAGAT 522
QY 540 VSLVFRQ---EDAFHPELMA---EPS----- 561
Db 523 VAVIVSRQOELAEQADQPEAKSAGVAVAPPAVPAAPAPPIPVOKSSSARSLEFTH 582
QY 562 ---QMOIPK-----ETKAEDEDIVLPDG--TREFLTFEYPLNDSSAGLGVSKG- 607
Db 583 QOQSOLNESQHTIDAGSEEAASNDSLPSSNSGMSHREELTHIPVHDTERKAGLGVSKG 642
QY 608 -----NRSKENHADLGIFVKSIINGAASKDKRLKVDOLIAVNGESLL 651
Db 643 TCSNLNASGSSASGSGMGLKHDGDLGIFVKVHIGGAASRDGRRLMNDQLSVNGVSLR 702
QY 652 GKTNDAMETTLRSM-STEGNKRGMQLVARRISC----- 687
Db 703 GONNAEMETTLRRAVNTTPGKHPTITLLVGRKILRSASSDILDSNSHSHSNGSGG 762
QY 688 -----NELSPGSPPELPIETALDDREIRISHLYSGIEGLDESPSRNALSR 738
Db 763 SMSNGSGNNSSNASDMSGATVYILSPERKEQKONGGGGSGACNEMNRKSNPLVDRL 822
QY 739 MG-----ESGKYQLSPVYVNMPO----- 755
Db 823 TGGICSSNSAOPSSQSHQOPHPQOQOORRLPAAPVCSAALNRESYMATNDNMP 882
QY 756 ---DVTYIIEDDRPLVLP-----PHLSDOSSSSSHD-----DVGTV 788
Db 883 AOMHLMTAHGNLTALLIEDAEPMSPTLPARPHDGOHNTSSANPSONLAVGNOGPPINTV 942
QY 789 TADACTWAKAALSDADCSLSPVDVPLAFOREGFGROGSEK-----RTK 834
Db 943 POTPTSSNFDATYSSQSLSTLN-SGVEHFSPDALCRKISIEKHHAALDARETGTYQRNK 1001
QY 835 QESDASQLD-EVYTKRS-----KSMDLGLAD-----ETKLN-T 865
Db 1002 KLREPERERRIQLTKSAVYGGISILFARIASANAQFGYKHAKTASSIEQRETQOLA 1061
QY 866 VDDKAGSPRDVPSLGLKSSSLESLOTAYAEVTLNDIPFHRPRPIIRGRCNESP 925
Db 1062 AAEARARQOLGDLPSLGLKSSSLESLOTWVQLOLMS-DEPRHQALRAPRGREGEDSL 1120
QY 926 RAAIDKSYD--KP-----AVDDDEGEMTLEEDTESSRSGRSVSTASQPSHSLEROM 979
Db 1121 RAAVSEPDASKPRKTWLEDD-----HEGGFASQRNG-----PFQSSLN 1161
QY 980 GNOEKGDYTKRKKDKTKGKKKRDOKKDKMAKKGMKLGDMFEGKHRRDKLEKTKG 1039
Db 1162 DCKH-GCKSSRAK-----KPSILRGIGHMFRFGKRRKDGVPVDN 1200
QY 1040 ---KIKIESFSEERIRMKOERIAQKTRFERROAREBDYAEIQDFHRTFGCDDDEL 1096

```

```

Db 1201 YAVNISPTISVSTATSPOLQOQOQ-----QOLQOHQOQO----- 1235
QY 1097 MYGVSYSEGSALNARPOSFREGHMDALYAQVKPRPNRSKSPVDNSNRSTPSN----- 1150
Db 1236 -----QIPTA--LAALERNGKRPAYQPPPLPAPNNGVSGNIGHTND 1275
QY 1151 -----HRIORLROEF--QAKKODEVDERRRTYSFEQPPWPNAP 1188
Db 1276 IFNHRYQHYANEDLHQOHQOHQISRRHQHYHSQARSQDV-SMHSSTSGSQPGSLAQP 1334
QY 1189 ATQS-GRHSYS-----VEYOMORQROBERESSQQAQROYSSLPRSRKNASSVSQDSWE- 1241
Db 1335 QASQDGVPRWSSYIYEYVQOQVGSIKSHSSA-----TSSSSPTINVPWMA 1384
QY 1242 ---QNTSPGEFQSAKENPRYSYOGSRNGYLGHGFMARVMEETOELLROE----- 1290
Db 1385 AAMNGYSPASLNSAR-----SRGPFV-----TQVTIREQSSGGIPAH 1422
QY 1291 ---ORRKEQOKKOPPS-----EGPSNYDS 1312
Db 1423 LQOHQOQOOLQOQOQPTYQTQVKMSGFSQYGS 1453

```

Search completed: July 24, 2002, 14:33:22
 Job time: 253 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:19:24 ; Search time 45.21 Seconds
(without alignments)
2882,042 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 7035
Sequence: 1 MKVTVCFGKTRRVVPCGDGH.....SQVARKLRLQTEPKGRPFYS 1356

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6332.5	90.0	1337	T13948	atypical protein k
2	1149.5	16.3	1464	T13716	bacopa gene prote
3	547	7.8	1360	T34302	cell polarity prot
4	361.5	5.1	2054	T46612	multiple PDZ domain p
5	352.5	5.0	2055	T30259	multiple PDZ domain p
6	335.5	4.8	2172	T20145	hypothetical prote
7	332	4.7	2466	T167629	protein tyrosine p
8	328.5	4.7	2450	T51625	protein-tyrosine-p
9	323	4.6	1131	T15617	hypothetical prote
10	321	4.6	2490	T454971	protein-tyrosine-p
11	314	4.5	2294	T167630	protein tyrosine p
12	296.5	4.2	5327	T13564	microtubule-associ
13	295	4.2	926	T138756	homolog of Drosoph
14	293	4.2	767	T09599	postsynaptic densi
15	293	4.2	1095	T43275	neurabin - rat
16	289.5	4.1	1410	T47137	hypothetical prote
17	286.5	4.1	904	T138757	homolog of Drosoph
18	286	4.1	724	TJH080	postsynaptic densi
19	286	4.1	870	T09458	channel associated
20	283.5	4.0	628	T09458	numb-binding prote
21	283.5	4.0	728	T09457	channel associated
22	279	4.0	852	T10811	synapse-associated
23	278.5	4.0	720	T45436	synapse-associated
24	278.5	4.0	911	T15652	brain-specific ang
25	278	4.0	1256	TJEO209	probable guanylate
26	272	3.9	1171	T42372	1-afadin - rat
27	260	3.7	1829	T41751	hypothetical prote
28	250	3.6	1012	T23160	s-afadin - rat
29	246	3.5	1663	T42092	

30	245.5	3.5	1553	2	T03301	rab3 effector prot
31	241.5	3.4	3488	2	T34418	hypothetical prote
32	240.5	3.4	1277	2	T14152	synaptic scaffold
33	239.5	3.4	2248	2	A35938	profilaggrin - hum
34	237.5	3.4	3924	2	S37431	ankyrin 2, neurona
35	228.5	3.2	1112	2	T32733	AMPA glutamate rec
36	228	3.2	2464	1	QRMSP1	microtubule-associ
37	227.5	3.2	960	1	A39651	discs-large tumor
38	226	3.2	2526	2	T20531	hypothetical prote
39	226	3.2	2722	2	T20532	hypothetical prote
40	226	3.2	2738	2	E88320	protein UNC-89 - C
41	226	3.2	6642	2	T29757	protein UNC-89 - C
42	225	3.2	754	1	UC5314	CDG28/cdc2-like ki
43	221.5	3.2	1281	2	T00346	hypothetical prote
44	221.5	3.1	771	1	A33430	h-caldesmon - chic
45	221.5	3.1	1666	2	T43169	hypothetical prote

ALIGNMENTS

RESULT 1	
T13948	atypical protein kinase C isotype-specific interacting protein ASIP - rat
C:Species: Rattus norvegicus (Norway rat)	
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000	
C:Accession: T13948	
R:Zumli, Y.; Hirose, T.; Tamai, Y.; Hirai, S.; Nagashima, Y.; Fujimoto, T.; Tabuse, Y.	
J:Cell Biol. 143, 95-106, 1998	
A:Title: An atypical PKC directly associates and colocalizes at the epithelial tight	
A:Reference number: Z17827; MUID:96437350	
A:Accession: T13948	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-1337 <ITD>	
A:Cross-references: EMBL:AB005549; NID:93868777; PIDN:BA43216.1; PID:93868778	
C:Genetics:	
A:Gene: asbp	
Query Match	90.0%; Score 6332.5; DB 2; Length 1337;
Best Local Similarity	89.7%; Pred. No. 7e-309;
Matches 1216; Conservative 63; Mismatches 58; Indels 19; Gaps 2;	
QY 1 MKVTVCFGKTRRVVPCGDGHMKVFSLIQAVTRKKAIAKDPNMYIOVRLHEHGDGIID 60	
DB 1 MKVTVCFGKTRRVVPCGDGHMKVFSLIQAVTRKKAIAKDPNMYIOVRLHEHGDGIID 60	
QY 61 LDDIICDVADDKRLVAVFDEODPHHGGDGTSSSTGTQSPETFGSELGTNNVSAFQPYQ 120	
DB 61 LDDIICDVADDKRLVAVFDEODPHHGGDGTSSSTGTQSPETFGSELGTNNVSAFQPYQ 120	
QY 121 ATSEIEVPSVIRANMPILHVRSSDPALIGLSTVSQSDNSFSEEPKRPMTWSTAGTL 180	
DB 121 ATSEIEVPSVIRANMPILHVRSSDPALIGLSTVSQSDNSFSEEPKRPMTWSTAGTL 180	
QY 181 KONTAGSPKTCRRKRDENRSLPRDTSNMSNOFORDNARSSLSASHPMVKMLKEQODE 240	
DB 181 KONTAGSPKTCRRKRDENRSLPRDTSNMSNOFORDNARSSLSASHPMVKMLKEQODE 240	
QY 241 DCTEEDNSKVEPVGADTGLEHIMPFSLDQMYKLVFVNDGSLGIHVVPFSAAGCGRTLG 300	
DB 241 DCTEEDNSKVEPVGADTGLEHIMPFSLDQMYKLVFVNDGSLGIHVVPFSAAGCGRTLG 300	
QY 301 LLYKRLKGGKAEHNLPRENDICIRINDGDLRNRFEQAOHMFQAMRTPLIMFHVVA 360	
DB 301 LLYKRLKGGKAEHNLPRENDICIRINDGDLRNRFEQAOHMFQAMRTPLIMFHVVA 360	
QY 361 ANKEQYEOLQSEKNNYSSRPSDQYIDNRSSVAGLHTQORAPRLNHPPEQIDSHSR 420	
DB 361 ANKEQYEOLQSEKNNYSSRPSDQYIDNRSSVAGLHTQORAPRLNHPPEQIDSHSR 420	
QY 421 LPHSAHPGSKRPASAPASAPQNVFTTVSSGYNTKIKIGKRLNIQLKKGTEGLCFSTISRQV 480	
DB 421 LPHSAHPGSKRPASAPASAPQNVFTTVSSGYNTKIKIGKRLNIQLKKGTEGLCFSTISRQV 480	

A:Reference number: 217708
A:Accession: T13716
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1454 <NC>
A:Cross-references: EMBL:AJ130871; NID:e1363519; PIDN:CAAI0224.1
C:Genetics:
A:Gene: bazooka
A:Cross-references: FlyBase:FBgn0000163
A:Map position: X

Query Match	Similarity	16.3%	Score	1149.17	48	Indels	537	Gaps
Best Local	Similarity	26.2%	Pred.	1.3e-49				
Matches	Conservative	212	Mismatches	470				
QY	1	MKYVCGFGRVYVPCGGGKMKVYSLQQAIVATYRRAIAKADPNYITQVHRLEHGDGILD	60					
Db	1	MKYVCGFGRVYVPCGGGKMKVYSLQQAIVATYRRAIAKADPNYITQVHRLEHGDGILD	58					
QY	61	LDLIDCDVADDDKRLVAFAFDEQD-----PHHGDDGTS-ASSTGTQSPETFGSELGTNNVS	114					
Db	59	PDDCVADVADDEQLAHFHDDPGDPGPVQGGGGAGSSSVSTGSPDIFRDP--TNEEA	116					
QY	115	AFQPYO-ATSEIEVTPS-----VLANNPLHYRHSRSDPALGLTSDVSDNFSESEPSRKN	169					
Db	117	PTCPRLDSTPHLEIVSTTSGPAAGLGVLAMKRSDDNLLA-SLAKEGSN-----	165					
QY	170	PTMSPTAGF-----LKQNTAGSPKTCDDKKDN-----YKSLPRTDSNMS	210					
Db	166	KRMSAARPYAGDSPELPLFDKKGQGLSPQWEEDDDPSHOLKEQLHQOQPHAAAGS	224					
QY	211	NOFORDNARSSLSASHPMVG-----KMLEKQF-QDEEDGTEE-----	245					
Db	225	SSGNHQPFAISGRSLSQMFLGDGNGYKMMEAAKLQNOPPAQYTQOQSHIAGGAGAVS	284					
QY	246	-----NSRVEPVGNADGLEHIFPFSLDDVAVKLVEPVNDGPGGLIHVVYPFSAARGRL	299					
Db	285	SKSLPRSKRKEPGLQAYESIRE-----KQDEMILLINEXGSPGLTALPDKHEGG--	335					
QY	300	GLIVYRLKEGKRAHENLFPENDCIYRINDGDLRRRREFQAOHNEROMAFRPIIFHVVYP	359					
Db	336	GLVQHVHPGSRARQGL-RRDRILLEINGIKLIGLTESQVQEOELRRLLE-----	384					
QY	360	AAANKQYEQQLSSKKNYSSRFSPDQYIDNRVNSAGLITVQAPRLNHPPEQIDSHS	419					
Db	385	-----SSLRKRVLRGDRNR-----RQGDQSKVADMEVATVS-----	417					
QY	420	RLPHSAHPSGKPPSAPASAPONVFSTTSSGYTKTKIGKRLNIOLKKGEGIGESITSRD	479					
Db	418	-----PLTRKHAAPVGTSLQY-----ATRRKIGKRIQIMLKKPGNGLGFVTTRD	462					
QY	480	VITGSGAPTYVNTLPRGAILQDRLKAGRLLEVNGVDLYCKRSQEVVSLRSTKMEGT	539					
Db	463	NPGAACHPIYINKLIPRGAILQDRLKAGRLLEVNGVDLYCKRSQEVVSLRSTKMEGT	522					
QY	540	VSLVLRQ-----EDAFHREINA-----EPS-----	561					
Db	523	VRLVSRQQLQELQADQPAEKSAAGVAVAPVAPVAAAAPPLIPVQKSSASSLFTH	582					
QY	562	-----OMQIPK-----ETKADEEDIVLTPPG--TRETFLEFVPLNDGSAGLGVSVKQ-	607					
Db	583	QOQSOLNESQHFIDAGESAASNDLSLPPSSMSHSRRETLHLPIYHDTKEKAGLVSVKQ	642					
QY	608	-----NRSKENHADLGIFVYSLINGGAASKDGRRLRYNQDLIYVNESILL	651					
Db	643	TCSULNASSGSSASSGNGMLKHDGDIQFVYKVIHNGGAASDGRRLRYNQDLIYVNGVSLR	702					
QY	652	GKTQADAMELURSM-STEGNKGMLQILVARIKSC-----	687					
Db	703	GONNAEMETLIRAWVNTPEKHPGTTLLVGRKILRSASSSDIIDHNSHSHSNGSG	762					
QY	688	-----NELKSPGSPPELPDLPETALDDREERRISHSLYSGIEGLDESPRMAALSR	738					

Db 763 SNSNGSNNNNSSNADSGATVYLYLSPEKREORNGCGGAGNENMRNSPVLDR 822
OY 739 MG-----ESGKYOLSPYVNMPO----- 755
Db 823 TGCCTSSNAGPSSQOOSHOOOPHPSOQOORRLPAAPVCSAALRNESYMATDNMSP 882
OY 756 -----DDVTIIEDRLPVLPR-----PHLSDOSSSSSHD-----DVGCV 788
Db 883 AQMHMLTAHGNLTALLIEDDAPMSPFLPARPHDGOHCNTSSANPSQNLAVNGOPIINTV 942
OY 789 TADACTWAKAALSDSADCSLPDVVYLAFOREGCRGMSERK-----PRK 834
Db 943 PGTPTSSNPDATYSSQLSLETN-SCVEHPSDALGRSISSEKHNAALDARETGYQRMK 1001
OY 835 QPSDASOLD-FYKTRKS-----KSMOLGIAD-----ETKLN-T 865
Db 1002 KIREPERERRIQLTKSANVYGCSTESLTRIASANAPSGYHNAKTASSIEORETIOOLA 1061
OY 866 VDDOKAGSPSPDYCPSLGLKSSSLESLOTAYAEVTLNGDIPFHRPRPIIRGCGNESF 925
Db 1062 AAEEAROLDGLDGLPSLGKSSSLESLOTAYAEVTLNGDIPFHRPRPIIRGCGNESF 1120
OY 926 RAADIKSYD-KP-----AVDDDEGMEILEEDTESSSGSGSVSTASDQPSHSLEROMN 979
Db 1121 RAAVSEPPASKPRKRTWLLDEGD-----HEGFPASQRMG-----PFGSSLN 1161
OY 980 GNOEGKDTDRKKDKTKGKKKKRDKERDKMKAKKGLGDMFPGFKRKDKLEKTK 1039
Db 1162 DGRH-GCKSSRAK-----KPSILGIGHMFPFGKRRKDGVPVYN 1200
OY 1040 ---KTKIOESTSEERIRMKOEIRIQAQTRFERQARERDYAETODFHRTFGCDEL 1096
Db 1201 YAVNISPPSTSVSTATSPQLQOQOQO-----QQLQOQOQO----- 1235
OY 1097 MNGVASYSGSALNARQSPREGHMDALYAQVKPRMSKSPVDSNSTPSN----- 1150
Db 1236 -----QIPYAA--LAALERNGKPPAYQPPPLPAPNGVSGNGIHND 1275
OY 1151 -----HRIQRLRQEF--QAKODEVEDRRRTYSFEQPMFNAPR 1188
Db 1276 IFNHRYQHANTEDLHQHQHQHISRRHQHYSQARSQDV-SMHSSTSGSPGSLADP 1334
OY 1189 ANQS-GRHSVS-----VEVOMORQROERESSQOAOQOYSSLPGRSKNASSVQDSWE- 1241
Db 1335 QAOQSDVRPMSSYEXEYVQOQOQVGSIKHSHSSA-----TSSSSSPINVHMKA 1384
OY 1242 ---QMSPEBGQSAKENRYSYSGSRNGYLGHGFMARVMELEQDLRQF----- 1290
Db 1385 AAMNGTSPASLNSSAR-----SRGPFV-----TQVTTIRESSGGIPAH 1422
OY 1291 --QRRKEQMKRQPPS-----EGSPNYS 1312
Db 1423 LLDQHQQOOLQOQOQPYQVQOKMSPQYGS 1453

RESULT 3
T34302
cell polarity protein par-3 - Caenorhabditis elegans
A:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34302
R:Bentley, D.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of C. elegans cosmid F54E7.
A:Reference number: 221502
A:Accession: T34302
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1360 <BEN>
A:Cross-references: EMBL:U00067; PIDN:AACT7513.1; GSPDB:GN00021; CESP:F54E7.3
A:Experimental source: strain Bristol N2; clone F54E7
C:Genetics:

A:Gene: par-3; CESP:F54E7.3
A:Map position: 3
A:Introns: 52/3; 102/1; 352/1; 411/1; 783/1; 832/1; 906/2; 990/3; 1049/2; 1265/2; 130

Query Match 7.8%; Score 547; DB 2; Length 1360;
Best Local Similarity 22.4%; Pred. No. 1.6e-19;
Matches 318; Conservative 170; Mismatches 444; Indels 490; Gaps 63;

Db 46 IOVHRLR-EGDGGILDDILDDICDVAD-DKDRVAVADE-----ODPHH----- 86
OY 22 IYHRLRLEASDGGILDDMDVLEEVFDLNTDQIATIDEANGSGTTPPYQIQOQHNYAQ 81
Db 87 -----GCDGTS-ASSYGT-----OSPFIQSELETNNVSAFOP----- 118
OY 82 PLPYARKFPGGSPSTPIASAFGSVTVNQHAAASPAFNVGPARNSMDFAPOFTHSKERD 141
Db 119 --YQATSEIEVTPSVYRANMPLHVRSSD-----PAL-IGLSTSVDSNFSSEEPS 166
OY 142 SVVEVSSFPQIQSGLRVSTPKPSRQSEVDYDCKPNNQPIRLRSLSLTEASGR--TEEAT 199
Db 167 RKNPTWSTTAGFLKONTAGSPKTCDRK--DENYRSLPDRDTSNMGNOFOROARSS--- 221
OY 200 PVKQSHVTLSPYEYKKLAODEKRSERKHYDN-----PGRARQSDRSRIT 248
OY 222 --LSASHPMVGKWLKOEODEDEDEEDNSRVEPVGHADTGLEHINPFSLDDMKLVEVYN 279
Db 249 DALLDARDRLADQLESQNAEE--TKSQMIRVK-----IDG--PMPTSLVTPPIPEKSE 301
OY 280 DGGPGLIHY-----VPFSARGRTIGLVLKLEKKGAKAHEMLFRENDCIVAINDG 330
Db 302 NEKQGLIEVNAVPEDESSELPSTSEPTKLSVQIMKIEDGRIKAKDRIR--VVRASIS 357
OY 331 DLNRRFEQAQHFQAMPTPIWFIHVYVPAKNEQYEQLSQSEKNYSSRFSPDQYID 390
Db 358 DL-----AAVTSRPVTLII 371
OY 391 NRSVNSAGLHTVQARPLNHPBQIDSHSLRPHASPSGKPPSAPASAPQNVSTTVSSG 450
Db 372 NRSLESF-----LEQES-----SAKPIQ--SALQQA----- 395
OY 451 YNFKTKGRLNIDQKKGTEGLSITSRDYITGGSAPYIYKNIPLPGAAIQDRLKAGDR 510
Db 396 -NTQYIGHYTVVELIKSSNGFPTVGR--TAKGERLFIYGYKPGVAL--GHLKSGDR 451
OY 511 LIEVNGVDLVGKQOEYVSLRSTKMEGYSLVFR--QEDAFHRELNAE----- 559
Db 452 LLETNGTPTGQMTQSEIVEKLEKTMGEEKIKFLVSRVSGSALINSTASSBNKENETLKY 511
OY 560 -----PSQM--QIPKETADEDDIVLPDGTREFLFEVPLNDGSGAGLGVSV 605
Db 512 VEDEKIPQKLPLPALMTPPVPKDTPA-----LSPGASRFEIVIPFINSSAGAGVSL 565
OY 606 KGNRSKENH---ADIGTFPKSIINGGAASKDGRLRVNDOLIAVNGESLCKTQMDAMEYL 662
Db 566 KAVRSKSSNGSKVDCGIFIKVMHGAAPKEGGLRVDDRIVGEDIDLEPLDREKQOAL 625
OY 663 RRGMSTEGKRGMIQILVARIKSCNELKSPG-----SPGPELPIETALDD 709
Db 626 AKKL-----KEVGMISNVRLTISRNEC--NPGQISRLDSLITVYDASSPSIS----- 671
OY 710 RERRISHLSYSGI-----EGLDSPSRNAALSRTMGSGKYYQLSPVNMPODDT 758
Db 672 -SRMSSHTAPDSSLPPPATRGTSAGSSGADSSHSRQSSAS----- 708
OY 759 VTIEDDLPLVLPRLHDQ-----SSSSSHDVGVVTADACTWAKAALSDSADCSLPDVP 814
Db 709 -----SAVPVPAARLTEROSIVSDGTSRND-----ESELPSDAD----- 742
OY 815 VLAFOREGFGROGMSERK-----TKQFSDASOLDVFKTRKSSMDLGIADETKLN 864
Db 743 --PFRNGGLGRKSLSKRGCAADQHIKLPD-----IKHQRONASAP----- 784

```

OY 865 TVDDQKAGSPRDVPSLGKSSLSLEOTATAVEYTLNGDIPFHRPRRIIRGNCNES 924
Db 785 TSSYOKR-----SKSPRSSONRYSPKLV-----DLP-----814
OY 925 FRAIDSYDKPAVDDDD-----ECMETLEEDTEESSRGRESVSTAS-----967
Db 815 TTAAASSTNGQNDSDMLNRRSQSMESINRPVESTLR-GTGOIPTGSSSKVOFMOAS 873
OY 968 -DQ-----PSHLEQRMNGNDEKDKTKRKKDKGKKRDKRDEKDKMKKGLGLGD 1022
Db 874 PDQHPFPGALLRLKN-----ESRSRDKSR-----RKGGMGNMRN 910
OY 1023 MFRFGKHKDKIKETGKIKIOESFTSEERIRMKQOEERIQAKTREFRERQARER-----1078
Db 911 FFGGSKSRDASPEKTPESVOLRSVERPKSLIDERNNSSERAPPLPPHQSQRSGSG 970
OY 1079 ---DYAEIOPHRTFGCDDDELMTGCVSYSGSM-----ALNARPOSREGHMMDA---1125
Db 971 NVFVDYGE-----PYGLIPQYPHNTTSGYESTADSELYDRYAAHRRHPRGPIIDEDY 1024
OY 1126 LYAQYKKPRNSKPSVDS-----NRSTPSNHDRIQRLRQEPQOAKODEDVEDRRRTYSPEQ 1181
Db 1025 ITRQOSTGN---SPINTSSYVNYGLPASN-----AYHVS 1057
OY 1182 PWPNAKPATOSGRHSVSYEVQOMOROEERESSQOAROYSLPROSKNNSVYSQDSWE 1241
Db 1058 RLP---PQTSNG---SISKTSGAMRRVYPAEYEDVAYHQ---QIPQOSTR-----1099
OY 1242 QVNSPGEFGQSAKENPRYSYOGSRNGYLGCHGNARVYLET 1283
Db 1100 ---YQGGSG---SGRGNDYHHMNSMFPATYGGAGVGAAPVKS 1137

RESULT 4
T46612
multi PDZ domain protein 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 21-Jul-2000
R/Accession: T46612
R/Unlmer, C.: Schumuck, K.; Figue, A.; Lubbert, H.
FEBS Lett. 424, 63-68, 1998
A:Title: Cloning and characterization of MUPP1, a novel PDZ domain protein.
A:Reference number: 223104; MUID:98196865
A:Accession: T46612
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2054 <URL>
A:Cross-references: EMBL:AF001320; NID:92959978; PIDN:CAA04681.1; PID:92959979
A:Experimental source: brain
C:Genetics:
A:Gene: MUPP1

```

```

Query Match 5.18; Score 361.5; DB 2; Length 2054;
Best Local Similarity 20.3%; Pred. No. 3; se-10;
Matches 246; Conservative 159; Mismatches 433; Indels 363; Gaps 48;

```

```

OY 189 KTCPRKK-----DEYRSIPRDTSNMSNOFQDNARSSLSASHPVQKMLEKQODEDGE 244
Db 3 ETIDKNRQLQAERLQSLKRGKRGVANDKRLSKVLYGS--PLFSQILSLQTSIQDLKD 60
OY 245 EDNSRVEPVGHADTGLEHIFNS-----267
Db 61 QVNVATLTATNADHA--HTPOFSSAIIISNLQSESLLLSPSNGNLLEAISGPGAPPAMDCKP 118
OY 268 ----LDDAVKL-----VEV-----PNDGPIGIHVVPFASARGRTGILVYKRLKSGKA 312
Db 119 ACEELDOIJKSMAQGRHVEIFELKPPCGG--LGSVYGLSKSNGELGIFVQELQESGVA 177
OY 313 EHENLFRENDCTIVRINDGDLNRNRFQQAQHFQAMRTPIIFHVVPANKEQYEQLSQS 372
Db 178 HRDGRKETEYDILAIN-----193

```

```

OY 373 EKNNTYSRSPDSQYIDNRSVNSAGLHTVORAPRLNHPEQIDSHRLSHSAHS-GKP 431
Db 194 -----GOVLDOTITHOAISTILOKA---KDTIQLVIAKSLPHISPRISRS 237
OY 432 PSAPASAPQNVSTTSSGYNFKKIGKRLNIQLKKGEGIGFSTISDYITGSAP--IYV 490
Db 238 PSA-----ASTVSAHSNPTHMOHVEITELVNDSGIGFEGI-----IGGATGVIV 282
OY 491 KNILPGRGAALQDGRKADRLIEVNGVDLKGSOEAVVSLSRKMGCTVSLVFR--QE 548
Db 283 KTIPLGVVADQHGRLSGDHILIKIDTDLAGMSSBQVAVR--QCNRYKLMARCAVE 340
OY 549 DAFHPREL-----NAEPSOMQIPREKAEDEDIYLPDGTREFLETPVPLNDSGSGIG 602
Db 341 EHPAPSSIGITLSSSTNISEMRVDSATQKNES-----ETPEVELT-KNVGLG 388
OY 603 VSYKGRSKENNADGIFPKSTIINGAASKDRLRVNQLAVNGESLIGTNDAMETL 662
Db 389 ITIAGYIGDKLEPSGIFPKSTIKSSAVELDGRIGQIVAVDGTNIGFTNGQAAVEVL 448
OY 663 RRMSTEGNKKRGMQILIVARR-ISKCNELKS--PGSPPELPETALDRERIRSHLSYS 720
Db 449 RHFGQI-----VRLTKRKGASQEAETTSREDTAKVDLPAENVEKD-----490
OY 721 GIEGLDESFSRNALSRINGESGKY-----OLSPYVNPQDDTVIIEDDRLL-----PYLP 770
Db 491 -----ESLSLKRSTISILPIEBEGYPLSTLEBEDYQQAALLTKMORIMGIVELIV 545
OY 771 PHLSDSSSSSHDDVGFVTADAGTAKAAISDADCSLSPVDVPLAFORBEFGQMSSE 830
Db 546 AHYSKSENGS-----LGISLEATVGHIFIRSVLPE-GPV-GHSGKLFSGDELLE 593
OY 831 KTRQFSDASOLDPVYTRKSKMDL-----GIADETKLNTVDDQKAGSPRDV 878
Db 594 VNGINILGENHODVYNLIKELPIDVTMCCRTVPTALSEVDSLIDHLELTERPHIDL 653
OY 879 GPSLGLKSSLSLEOTATAVEYTLNGDIPFHRPRRIIRG-----RGNCNES---F 925
Db 654 GEFIG---SSETEDPMLAMDYQNAE-ETQPLAMWAGIAQIELEKSGRIGFSILDY 709
OY 926 RAADKSYDKPA-----VDDDEGMEULEEDTE--E 954
Db 710 QDPID-----PANTVIVIRSLVPGIAGKDGRLPEGDRIMFVNDIENSTLEAVEALK 764
OY 955 SSRGREGSVTASQPSHLSEROMNGOEKGDTRKKDKTCKEKKDKRDEKDKMKARK 1014
Db 765 GAPSQWRTIGYAKPLPLSPBGIYSAKED---TFLCSPTCKE-----804
OY 1015 GMLKGLGDMFFRGHRKD-----DKTEKTKIKIQESFTSEERIRMKQOEERIQAKTR 1068
Db 805 ---MGLSDKALF---RADLALIDTPDAESVAESRFEQSPDNISYSTO-----848
OY 1069 EFRERQARERDYAEIOPHRTFGCDDDELMTGCVSYSGSMALNARPOSREG-----1120
Db 849 -----ASVLSLH-DGACSDGMNTG-----PLPSSPPKPDVYTNSSDLV 884
OY 1121 ---HM-MDALVQO---VKPRNSKP---SPVDSNRSTPSNHDRIQRLRQEPQAK-----Q 1166
Db 885 LGLHLSLEELVYQNLQOHAGSPPTDMSPATSGFTVSDTPPANAQKRECAVTAMT 944
OY 1167 DEDVEDRRRTYSFEQPMNARP---ATQSGRHSVVEVQOMOROEERESSQOAROYSS 1223
Db 945 PSQLSGLSTLTPALPALPAVAPKYLTEOSSILVSDAESVTILOSQDEAFERTVITAKSSS 1004
OY 1224 L 1224
Db 1005 L 1005

RESULT 5
multiple PDZ domain protein - mouse
T30259
C:Species: Mus musculus (house mouse)

```


Db 1648 PANTSNTWSSALHQTLSNMWQASHHEAPKQSEDITCTMYTPQKIPNKEEDSNPS 1707
QY 807 SLSPDVPLAFOREGFGROSMS 829
Db 1708 PLPPDMAFGOSYOPQSESASSS 1730

RESULT 8
S71625
protein-tyrosine-phosphatase (EC 3.1.3.48) RIP - mouse
N:Alternate names: epidermal growth factor-binding protein; serine proteinase
C:Species: Mus musculus (house mouse)
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
C:Accession: S71625; S67987
F:Child: D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe, T.
FEBS Lett. 358, 233-239, 1995
A>Title: Characterization of a protein tyrosine phosphatase (Rip) expressed at a very ee
A:Reference number: S71625; M01D:95145716
A:Accession: S71625
A:Molecule type: mRNA
A:Residues: 1-2450 <CHI>
A:Cross-references: EMBL:D83966; NID:g1232103; PDB:BAAL2158.1; PID:g1232104
R:MolF: B.B.; Brown, M.D. 1995
FEBS Lett. 376, 177-180, 1995
A>Title: Epidermal growth factor-binding protein activates soluble and receptor-bound s
A:Reference number: S67987; M01D:96105375
A:Accession: S67987
A:Molecule type: protein
A:Residues: 1098-1102 <MOL>
A:Experimental source: submaxillary glands
C:Genetics:
A:Gene: Pcpn13
A:Map position: 5
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology;
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; tyros
F:566-860/Domain: protein 4.1 membrane-binding domain homology <B41>
F:1089-1165/Domain: GLGF domain homology <GLG1>
F:1361-1437/Domain: GLGF domain homology <GLG2>
F:1495-1574/Domain: GLGF domain homology <GLG3>
F:1769-1840/Domain: GLGF domain homology <GLG4>
F:1863-1937/Domain: GLGF domain homology <GLG5>
F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PPI1>
F:2274/Active site: Cys (phosphocysteine intermediate) #status predicted
F:2380/Binding site: substrate phosphate (Arg) #status predicted

Query Match 4.78; Score 328.5; DB 2; Length 2450;
Best Local Similarity 19.84; Pred. No. 3.2e-08;
Matches 270; Conservative 167; Mismatches 506; Indels 423; Gaps 55;

QY 90 GTSASSTGTO-----SPEIFSELGTNNVSAFOYQATSELEVPPSVLRAMPPLHVR 141
Db 902 GSLLASSTINKLAVRPLSVQAEIL-KRLSSSEWSLYQPLDNSSK-EKIDKASWEEKPR--- 956
QY 142 RSSDPALIGLSTVSDFSSESEPSRKNPTKWTAGFLKONTA---GSPPTCDKKDEN 198
Db 957 -----GMSKYHDLQSQSLCPHRKQVIMEA---LPQAFALVIGKPL----- 995
QY 199 YRSLPRTSMKSNQFOFQDNARSSLSASHPMVGKLEKQEDQEDQEDQEDNSKVEEYVQ---- 254
Db 996 YPMARSTESLAGLPKLDNSSVSA-----LNRSPERRNHESDSTEDPCQAYVVGMSLP 1050
QY 255 -----HADGLEGHINPESLDVWVKLEVNDGPGIGIHVPPSARGR 297
Db 1051 SSGKSSSQVPPKNDTJHKRSIVSSPEREI-TLVNLKKDPKHG--LGFQITIGGEKMGRL 1107
QY 298 TLGLLVRLEKGAKEHNLFRENDCIVRINDGLJRNKRFPQAOHMFQAMR----- 349
Db 1108 DLGVFISAVTPGCPADLDGLCKPGRDLISVSVSLGEGSHHAADVILQNAPEDEVTLVTSQ 1167
QY 350 -----TPIWTFHVVPANKE----- 364

Db 1168 PKERPSKVPSTFVHEFANCMKYTKKPAVMODSAMPDSEDQWPFRGTLRIHIESPGLSG 1227
QY 365 -----OYEOLSOSEKNNYSSRF-----SPDQYIDNRVNSAG 398
Db 1228 LREGSLSDQSDRTESASISQSQVQNGCFPASHLDKRWQEPQSHSPSPSTYTKVNETFSDS 1287
QY 399 LHTVQARPLNHPPEQID-SHSRLPHSAHPSGKPPSAPASAPONYESTTVSSGYNTK--- 454
Db 1288 NRSKAKRRCISDLEHLHLCADSDKDDSTYTSQDQHTSKOEPESSLSSTSNKTSPTSSAS 1347
QY 455 --KIGKRLNLOKKCTBGLFSPIT--SRDYTTGSGADYVKNILIPRAA10DGRLLACD 509
Db 1348 PPKGDDFEVELAKIDGSLGISYTGVTGYNHG---IYKAA11PKGAASDGRIRHGD 1404
QY 510 RLIEVNGDVLGKSGOEVSILRSTKMGVTSLV-----FREDQAFHPELNA-EPS 561
Db 1405 RVLAVNGVSLGATHKKQAVELRLNTGQ--VHLLLEKQVPTSRQDPAGQSPPPDDA 1462
QY 562 QMOIP---KETKAEDEDIVLPDGTREFLTFEYPL-NDSGSAGLVSVKGRSKENHAD 616
Db 1463 QROAPEKVAQTPHYKQVSPYTEDN-----TFEVKLFKNSSGLGFSFREDNLIPQJNG 1517
QY 617 LGIFVKSIIINGASKDRKLVNDQLIYNGESLGTNODAMETLRSMSTEGKRGKI 676
Db 1518 SIVRYKKLFPQOPAAESKIDVGDVILKNGAPLKGLSQODV1SALRGTAPE-----V 1570
QY 677 QLIYAR-----RIKCNELKSPGS-----PPGELPIPTALDDRRRISLSLY 719
Db 1571 SLLICRPAPGVLPEDITFLNPLSPANSFLNKSSETQSPSSVQGS-----SHD-- 1622
QY 720 SGIEGLDE---SPSRNALSRIMGESGRYOLSPYVNNPODDVYIIEDDRPLVLPRLHS- 774
Db 1623 NGVSGKTKNHCRAPRSRYSYSD-HSESG-----EDDSV-----RAPAKPNVTR 1665
QY 775 -----DQSSSSSHDY-----GFV1ADACTWAKAISDSADCSLSDPDVPLAF 818
Db 1666 VAAPRHEAPRQOESICAMFYLPKRIPLKLESESHPPPLDVSPOQOPPAECAP---- 1721
QY 819 QREGGRQSMSEKRRKQSD-AQSLDFVTKRSKSMIDG-----IADETRLNVDDOK 870
Db 1722 -----SDAGKHFTHLASQLSKEENTLTLNDJGNHLEDELEVELLITLVKSEK 1771
QY 871 A-----GSPS-----RDV--GPSLIG---LKSSSLESL-QTVAEYTLNGDIPFHRP 911
Db 1772 GSLGFTVTKGSGSIGCYVHDVYQDPKAGDGRLLKAGDRLIKVNDTVMTHTDANLLRA 1831
QY 912 RPRITR-----GRG-----CNDS 924
Db 1832 APKTVRLVGLRIELPRMPVPEPHLLPDITVYCHGEELGFSLSGGGSPHGVVYISDINPR 1891
QY 925 FRAAIDKSYDRPAVDDDEGMEETLEEDTEESRS-----GRESVST--A 966
Db 1892 SAAAVDGLQDLIDIIHYVNGVSTQGTLEDANRALDLSPSVLYKVTYDGCYVFTTAA 1951
QY 967 SDPSSHLEROMNGNQEGKDTDRKKDTGKEKKDKRDKKDKAKKAKMLKGLG---DM 1023
Db 1952 ISAPRFTKANGLTSMPEGQPALMKNSFSK-----VNGEGVHEAV 1992
QY 1024 FRFGKHKKDDKIEKTKIKIQESTSEERIRKMOEQERIA-----KTRERERQA 1075
Db 1993 CPAGESSSQMKESAGLIEKTESNSRDDIYDDPEAEV1OSLIDVVEAOONLLNORA 2052
QY 1076 RERDVAE1ODPHRFPGCDEELMVGVSVEGSMALNARQSPREGHMMDALYQAQKPRN 1135
Db 2053 TRR-----ACSPD-----PLRTGEAPEEG---DTDY-----N 2077
QY 1136 SKPSPVDSNRSTPSNHRIO-----RLROEQOAKQODEVEDRRT 1176
Db 2078 GSP1PEDVPEVSQEGKVLDLASLTAASQOEKPIEDATQESRNET 2123

RESULT 9

t15617

hypothetical protein C25F6.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

R:Benley, D.
 Submitted to the EMBL data library, October 1995
 A:Description: The sequence of *C. elegans* cosmid C25F6.

A:Reference number: Z18377
 A:Accession: F113617
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-1131 <BEN>
 A:Cross-references: EMBL:U39742; NID:g1049455; PID:g1049459; PIDN:AAA0434.1; CESP:C25F6

A:Gene: CESP:C25F6.2
 A:Introns: 105/1; 210/3; 283/3; 316/1; 346/2; 463/3; 566/1; 722/2; 897/1; 991/2; 1032/1;

Query Match 4.6%; Score 323; DB 2; Length 1131;
 Best Local Similarity 22.5%; Pred. No. 2e-08;
 Matches 176; Conservative 105; Mismatches 304; Indels 198; Gaps 29;

Db 817 GSRSSLSFSKRPKPPK-----FVKSTDRL-----NDLNESSVAEEPV-----WSYQAVE 859
 QY 802 DSA 804
 Db 860 QQA 862

RESULT 10
 A54971

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 13 - human
 N:Alternate names: Fas-associated phosphatase FAP-1; protein-tyrosine-phosphatase hpr
 C:Species: *Homo sapiens* (man)
 C:Date: 11-Nov-1994 #sequence_revision 08-Feb-1996 #text_change 01-Dec-2000
 A:Accession: A54971; A55114; I55955; I53483; S46955
 R:Barville, D.; Ahmad, S.; Stocco, R.; Shen, S.H.
 J. Biol. Chem. 269, 22320-22327, 1994

A:Title: A novel protein-tyrosine phosphatase with homology to both the cytoskeletal
 A:Reference number: A54971; MUID:94350588
 A:Accession: A54971

A:Molecule type: mRNA
 A:Residues: 1-2490 <BAP>
 A:Cross-references: GB:U012128
 R:Saras, J.; Claesson-Welsh, L.; Heldin, C.H.; Gonsky, L.V.
 J. Biol. Chem. 269, 24082-24089, 1994

A:Title: Cloning and characterization of PTP11, a protein tyrosine phosphatase with s
 A:Reference number: A55114; MUID:95014139
 A:Accession: A55114

A:Molecule type: mRNA
 A:Residues: 1-61; 'GS', 64-839, 'D', 841-1055, 1075-1133, 'FH', 1136-1210, 'T', 1212-1383, 1389
 A:Cross-references: GB:X60289; NID:9515030; PIDN:CA56563.1; PID:9515031
 R:Arato, T.; Tyle, S.; Kitada, S.; Reed, J.C.
 Science 268, 411-415, 1995

A:Title: FAP-1: a protein tyrosine phosphatase that associates with Fas.
 A:Reference number: I59595; MUID:95323528
 A:Accession: I59595

A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1279-1888 <RES>

A:Cross-references: GB:I34583; NID:9806291; PIDN:AA041755.1; PID:9806292
 R:MacKawa, K.; Imagawa, N.; Nagamatsu, M.; Harada, S.
 FEBS Lett. 337, 200-206, 1994

A:Title: Molecular cloning of a novel protein-tyrosine phosphatase containing a membr
 A:Reference number: I53483; MUID:94116679
 A:Accession: I53483

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-61; 'GS', 64-839, 'D', 841-1210, 'T', 1212-1383, 1389-2299, 'QW', 2302-2490 <RE2
 A:Cross-references: GB:U012128; NID:9452189; PIDN:BA04750.1; PID:9452190

C:Gene: GDB:PTPN13
 A:Cross-references: GDB:306348; OMIM:600267
 A:Map position: 4q21.3-4q21.3

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GIGF domain homolog
 F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>
 F:1099-1175/Domain: GIGF domain homology <GIG1>
 F:1373-1454/Domain: GIGF domain homology <GIG2>
 F:1511-1590/Domain: GIGF domain homology <GIG3>
 F:1799-1870/Domain: GIGF domain homology <GIG4>
 F:1893-1967/Domain: GIGF domain homology <GIG5>
 F:2242-2461/Domain: protein-tyrosine-phosphatase homology <PPT1>
 F:2413/Active site: Cys (phosphotyrosine intermediate) #status predicted

F:2419/Binding site: substrate phosphate (Arg) #status predicted

Query Match 4.6%; Score 321; DB 1; Length 2490;
 Best Local Similarity 20.9%; Pred. No. 7.7e-08;
 Matches 196; Conservative 128; Mismatches 364; Indels 250; Gaps 32;

Db 80 DEOPHNGGCTGASSTGQSPETFGSELG-----TNNVSAFOPOATSER--E 126
 Db 879 DAQDI-----ERASFSLNIDQAESVKGFMNGRAITGSLASTLNKLAIVKPLVSVAQAEIKR 934

QY 743 G-KYOLSPVNNPQDDYIIIEDRLPVLPPLHSDQSSSSSHDDVGFVTADAGTAAATIS 801
 Db 757 DEWMTARKVHENGEEETAGVLPSSKRVKRRRLKROVNFNSGSSISLGRNNSSTTGLENR 816
 QY 709 DR---ERRISHSLYSIGIEGL-----DESSRNAALSRIMGES 742
 Db 757 DEWMTARKVHENGEEETAGVLPSSKRVKRRRLKROVNFNSGSSISLGRNNSSTTGLENR 816
 QY 743 G-KYOLSPVNNPQDDYIIIEDRLPVLPPLHSDQSSSSSHDDVGFVTADAGTAAATIS 801

```

OY 127 VTPSVLRANMPH--VRSSSDPAL-----IGLSTVSDSNFSSEPSRRKNPTRMSTTAGF 179
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 935 LSCSELSTLQPLQNSKKEKNDKASWEEKPREMSKSYHDLQSALYHRRKN----- 984
OY 180 LKQNTAGSKTCDRKDEKDEYRSLPR-DTSMNSQKQFORDNARSSLASHPMVKMLEKQDQ 238
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 985 VLVNNEPPPTVAELVKGPSHQMSRSDASLAGVYKLNNKSVASLN-----RSP 1034
OY 239 DDGTFEEDNSRVEPVGHA-----DTGLEHLPNFSLDLM-----VKLV 275
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1035 ERKKEHSDSSSTIEDQOAVYLGKTHMSSGNSSQVPLKENDVLHKWSTVSSPERETITLV 1094
OY 276 EYVNDGG-PLGIHVYFVSARGRTIGLLVRLKLEKGAHEMLFRENDCIVRLNDGDLRN 334
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1095 NLKKAQYGLGFOITGGEKMGRLDIGIFISSVAPGSPADLDGLKPGDRLISVNSVLBG 1154
OY 335 RFEQAOHMFROA-----MRTPIWHVYPA-----NKQYED----- 368
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1155 VSHHAATEILQNAPEDEVTLVISQPKERISKVPSTPVHLNEMKNYMKSSYMDSDAMSS 1214
OY 369 -----LSQSEKNYYSRFR----- 382
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1215 SKDHHSKGTLLHISENSFGSGGLREGSLSDSODRTESASLSQSOVNGFFASHLGDQTM 1274
OY 383 -----SPDSQYIDNRSVNSAGLHTVQ---RAPRLNHPEPID-SHSRLPHSAHSGKBP 432
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1275 QRSQHSFSPSVYSKATEKETFTDSNOSKTKKPGISDVTDYSDRGSDMEATYSSSDQH 1334
OY 433 SAPASAPQVNSTVSSGNT-----KTKGRNLQOLKKGTEGLGFSIT-----SRD 479
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1335 QTPKQESSSVNTSMKNKFTSSSPKPGDIFEVELAKNDLSGISVYLEDKGGVNTS 1394
OY 480 VTIGSAPITVYKNILPRGAIDGRLKAGDRLIEVNGVDLVKSGQEEVYSLRSTMEGT 539
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1395 VNHGG---IYKAVITPGAAESDGRHKGDRLVAVNGVSELEGTHQAVETLANT---GQ 1448
OY 540 VSLVLFRODAFHPR-----LNAEPSQMLP---KETKADEDIVLTPDGTRE 586
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1449 VVHLLEKQSPKSEHVPTQCTLSQNAQOGGPEKVKTTQVYDSFV-----TER 1502
OY 587 LFEFVPL-NDSSAGLGVSVKGRSKENHADLGIFKSLIINGAASKDGLRYNDOLIV 645
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1503 MTFEYKLFENSSGLGFSFRENLLPEQINASTIVRYKKLFPGQPAESKIDVGLVLY 1562
OY 646 NCESLGKTNODAMETLRRSMSTEGNKGMIQILVARRISKCNELKSPSPGPELPIET 705
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1563 NGASLKLGSQEVISALRGTAPE-----VFLLCR-----PPPVLEPIDT 1603
OY 706 ALDNERRISHSLYSGIEGLDESFSRNALSLRI-----WGESGKYQL-SPTVNNP 754
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1604 AL-----LTFLOSPAQVLPNNSKDSQSCVEOSTSDENEMSDKSKQCKSPSRDS 1656
OY 755 QDDVTIIEDRLPYLPHLSDOS-SSSHDDVGFVYADAGTAKAALIS----- 801
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1657 YSDSSGSGEDDLVTAANLSTNSWALHQTLSNMVSOASHHAPKSOEDTCTMFTYFP 1716
OY 802 -----DSADCSLSPVDVPLVAFORCGFGRGMS 829
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1717 QKIPNKEFEDSNPSPLPDMAPGOSYQPSSESASSS 1754

```

```

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2294 <RES>
A:Cross-references: GB:D21211; NID:9452193; PTDN:BA04752.1; PTD:9452194
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homolog
F:574-868/Domain: protein 4.1 membrane binding domain homolog <B41>
F:1182-1258/Domain: GLGF domain homology <GLG2>
F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match 4.5%; Score 314; DB 2; Length 2294;
Best local Similarity 20.3%; Pred.No.1.5e-07;
Matches 201; Conservative 143; Mismatches 360; Indels 288; Gaps 38;

OY 25 SLIQAVTRYRKAIADPNYMIQVHRLHEDGSLDLDLTL-----CDVADKDRIVAYV 79
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 668 SLIQHTLTCQ-----YYLDLRK-----DLBERMNC-DKSLSLASL 705
OY 80 DEODPHHGDDTGSASTGSPETFGSELGTNNVSAFO-----PYQATSEIYTPSVLRA 134
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 706 ALQ-----AEYGDYQPEVHG-----VSFMRMHHYLPARMKELDL-SYIKE 745
OY 135 NMP-LHYRRSSDPALIGLSTVSD-----SNFSEEPSRRKNPT---RKSTT 176
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 746 ELPRKLH-----NTYVGASEKETELFELKVCQRLTEYGVHFNHVPHEKKSQGTILLGVCS 799
OY 177 AGFLKONTAGSPKTC-----DRKK---DENYRSLPRDSNMNSNOFORNARSSLASH 226
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 800 KGVLYEYVNHGVRTVLRFWRRETKKISFSKKKITLQNTSDGKJHGFQNTSKICQYLH 859
OY 227 -----PVMKMLEKQDDEDEGTEEDNSRVEPVGADTGLEHLPNFSLDLMKLYE 276
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 860 LCSYQHKFOLQMAROSNOADIDVLHKRMSIYS-----SPREITLVN 904
OY 277 VPRDGG-PLGIHVYFVSARGRTIGLLVRLKLEKGAHEMLFRENDCIVRLNDGDLRN 335
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 905 LKRDATYGLGFOITGGEKMGRLDIGIFISSVAPGSPADLDGLKPGDRLISVNSVSEGV 964
OY 336 RFEQAOHMFROA-----MRTPIWHVY-----VPAAN 362
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 965 SHHAATEILQNAPEDEVTLVISQPKERISKVPSTPVHLNEMKNYMKSSYMDSDAMSS 1024
OY 363 KEQY-----EOLSOSEKNYYSRFR----- 382
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1025 KDHHSKGTLLHISENSFGSGGLREGSLSDSODRTESASLSQSOVNGFFASHLGDQTM 1084
OY 383 -----SPDSQYIDNRSVNSAGLHTVQ---RAPRLNHPEPID-SHSRLPHSAHSGKBP 433
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1085 ESOHGSFSPSVYSKATEKETFTDSNOSKTKKPGISDVTDYSDRGSDMEATYSSSDQH 1144
OY 434 APASAPQVNSTVSSGNT-----KTKGRNLQOLKKGTEGLGFSIT---SRDVTIGS 485
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1145 TPKQESSSVNTSMKNKFTSSSPKPGDIFEVELAKNDLSGISVYLEDKGGVNTS 1203
OY 486 APIYVKNILPRGAIDGRLKAGDRLIEVNGVDLVKSGQEEVYSLRSTMEGTSLIV 545
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1204 --IYKAVITPGAAESDGRHKGDRLVAVNGVSELEGTHQAVETLANT---QGVALL 1258
OY 546 RQDADFHPR-----LNAEPSQMLP---KETKADEDIVLTPDGTREPIFEV 592
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1259 EKQSPKSEHVPTQCTLSQNAQOGGPEKVKTTQVYDSFV-----TEENFEVK 1312
OY 593 L-NDSSAGLGVSVKGRSKENHADLGIFKSLIINGAASKDGLRYNDOLIVNGESL 651
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1313 LFNSSGGLGFSFRENLLPEQINASTIVRYKKLFPGQPAESKIDVGLVLYKNGASLK 1372
OY 652 GKTNODAMETLRRSMSTEGNKGMIQILVARRISKCNELKSPSPGPELPIETALDRE 711
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1373 GLSQQEVISALRGTAPE-----VFLLCR-----PPGVLEPIDTAL---- 1409
OY 712 RRISSHLYSGIEGLDESFSRNALSLRI-----WGESGKYQL-SPTVNNPQDDTVI 760
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1410 ---LTFLOSPAQVLPNNSKDSQSCVEOSTSDENEMSDKSKQCKSPSRDSYSDSSG 1466

```

QY 761 IEDRLPVLPHLSDQS-SSSHDDVGFVTADAGTWAKAALIS----- 801
 Db 1467 SGEIDLVTAPANISNTSMSSALHQTLSNMVSOQASHHEAFKSOEDTICTMYYPQKIPNK 1526
 QY 802 ----DSADCSLSPDYDVLAFAFGREGROGWS 829
 Db 1527 PEPEDSNPSPPLPDMAGQSYOPQSBASASS 1558

RESULT 12

T13564
 microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
 N:Alternate names: hypothetical protein EG:49E4.1
 C:Species: Drosophila melanogaster
 C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13564
 R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
 Submitted to the EMBL Data Library, April 1999
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A:Reference number: 217689
 A:Accession: T13564
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-5327 <SPA>
 A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0025392
 A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A>Note: EG:49E4.1
 C:Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match 4.2%; Score 296.5; DB 2; Length 5327;
 Best Local Similarity 18.6%; Pred. No. 3.8e-06;
 Matches 279; Conservative 227; Mismatches 569; Indels 423; Gaps 60;

QY 91 TSASSTGOSPEIFGSLGNNVSAFOYQANSEIVTSPVLK-----ANMPLHYRRSS 144
 Db 693 SSPESTAKAK-----EANNKKVLESKOQARVQATSTSRVYTSASERVOQQAFA 746
 QY 145 DPALIGLSTVSQDSNFSSEPRKNTPTKMTAGF-LKONTAGSKTCDRK----- 195
 Db 747 KITATG-AVQATQRPKIPSRPGVSPSKRAPPGSPVNO-AKKRADLKKTRLDKGT 802
 QY 196 -DENYSLP-RDTSNMSNFPORDNARSSLSASHPVAGKMLEKQEDDEGTEDN-SRVE 251
 Db 803 TDSLSLVSTSPADATAAKLDLTASOELDA-----EKORLIDLKEQEVAREIE 853
 QY 252 PVGHADTGLEHLPNFSLDLM-----VKLVEYPNCGPLGIHVVPFSARGRTGLL 302
 Db 854 AV-----FSRDEMKGROHQOIKALREMPAGETGDG----- 884
 QY 303 VKRLKGGKAHEHNFNDQIVIRINDGLNNRRPEQAOHMFROAMPTILWFHYVPAAN 362
 Db 885 -----ENEDDEBEYLITKEVEEQYTEDSI-----VFOESSM----- 918
 QY 363 KEQYEQLSQSEKNNYSSRFSPDSQYID-----NREVSNAGLHTVORADRLNHPPE 413
 Db 919 KEEELQKHORSQSEKKRKSAEELEALAYEAERKARLEGASARDESELDVEPE 978
 QY 414 QID-----SHSRLPHS-AHPSGKPPSAPA----- 436
 Db 979 QSKIKAEVODIATAKDIASRTTEOLAKPAEELSSPPEEKLSTKTDITDDQIGAV 1038
 QY 437 -----SAPQNVSTTVSSGNT-----KKIKRLNIO---LKKGTESIG 472
 Db 1039 DVLFPVNLQESLPREKFSATIESGATTAPLPEDERIPLDQIKEDLVIEKVKETKEAE 1098
 QY 473 FSTSRDVTIGGSAPYIVKNILPRGA-AIDGRUKA-----GDRLI-----EVN 515
 Db 1099 AIVVAVYQVTLPEAAPIADITLILASATKADAPKDANAEALGELPDSERVLPMKMTFEAQON 1158

QY 516 GYDLVGCSEDEV-----VSLRSTKMEGTSLVFERODAPHRR----- 554
 Db 1159 LLRVITKTPDEVDLPVHEADLGLYKDSQDANKSISHKESKKEKFTDDEKENVG 1218
 QY 555 -ELMPEQSQMIP-----KETKAEDDIYLTDPG-----REFLFEVPLNDGSGAGLV 603
 Db 1219 EIELGDBPKNVDSHVLKLESVOEVAEKVYLETTEVFERKOEIVAEATVITQENQEDLME 1278
 QY 604 SVKGNRSKENHADLGIYKVSIIINGAASKDR-----LRVNDQILA-----VAGESLLG 652
 Db 1279 QYVKDEEHQKTESGITTEKEAKKASPTPEEKESDIDSDELTAQOLADPTVPPKSAKD 1338
 QY 653 KTNODAME-----TLRRSMTEGKKRMIDLIYARISCKNELKSPGPELPTETALDD 709
 Db 1339 REDTGSIESPTIEAIEVE-----VOAKOEAKRP--VPAPDEAIKT----- 1378
 QY 710 RRRRIISHLTSGIEGLDESPSRNALSRINGESG-----KYOLSPVNNPQDDTVY 760
 Db 1379 -----EKSPILASKETSPESATGVSVEDTEOTKSKSPVSPPESEA-- 1420
 QY 761 IEDRLPVLPHLSDQS-----SSSHDDVGFVTADAGTWAKAALISDSADCSL 808
 Db 1421 -KDKSPFASGEASRPESVASEVDEAGKAEISRESIAKTHKDESSLDAKAKQESRRESL 1479
 QY 809 SPDDVPVLAFAFGREGROGWSMEKTKQFSDASOLDPVKTRKSK-SMDGLIADETKLVTV 866
 Db 1480 AESIKP-----BSGIDEKSAKASKASRPESV-TDKSEKSPRESIESLSKAST 1528
 QY 867 DDQKAGSPSRDVG-----PSLGLKSSLSLELQTA 896
 Db 1529 KDKSAPPKSEKASRPESVASEVDETEKSEKSPRESIASAKPPIEFREVSREPISIDG 1588
 QY 897 VAEVTLNGDLPTRPRRIIRGRCNESFR-AAIDKSYDKPAVODDDEGNETLEOTEE 954
 Db 1589 IKDESA-----KPSRKSDSPASKASRPESVLESVKDEIKETKESRESVAESFKA 1641
 QY 955 SRSRGRESVSTASD--QPSHSLEROMNGQKGDITDKKDKTKGKRRDKKEDKMKKA 1012
 Db 1642 DSTKDEKSLTISKDISRPSAENVMDAPFK---ETSPRESAVSGMKDESKSPSRRES 1698
 QY 1013 KR-----GMLKGLGDMFRFGKHKDKDKIKTKIKITIOESTSEERIR 1055
 Db 1699 VKDGAQSQRETSRPASVASEAKDADLKLRSRPSITTSQSEACISIDKESPLASEPESR 1758
 QY 1056 MKQEOERIAQKRREREQAREROYAEIQDHRFGGDDDELMTGGVSSYEGSMALNAPQ 1115
 Db 1759 PASVAESVKDEAKESKESRRE-SVAE-----KSPILPSKEAS-----RRA 1797
 QY 1116 SPREGHMADALYAQVKRRNS--KPSFVDSNR-STPSNHDRIQLRQEOQAKODEVD 1172
 Db 1798 SVAESIKDEAKESKESRRESVAESKSPILPSKEASRPAS--VAESINDAEKESK-----ES 1851
 QY 1173 RRRRTYFPOPUPN---RRPATQSGRHSVSEVOMQRODERESSQOAROYSSLPSPROSR 1229
 Db 1852 RRESVAESKSPILPSKEASRPAS--SVAESIKDEAKESKESRRESVAESKSPILPSKA 1904
 QY 1220 KNAASVSO---DSWEQNYSPGEGFOSAKENPRYSYOGSGNCTYLGGHGNAMVLETOL 1286
 Db 1905 SRPASVAESIKDEAKESKESRRESVAESKSP-LPSKEASRPASV-----VASINDAEK 1957
 QY 1287 LRQORRKEQOMKOPPSRPSNYDSY-KKVQDPTAPPKGPFRODV--PPSPSOVA 1340
 Db 1958 SKESRRESVAESKSPILPSKEASRPASVASEIKDEAKESKESRRESVAESKSPILPSKA 2015

RESULT 13

138756
 homolog of Drosophila discs large protein, isoform 2 - human
 C:Species: Homo sapiens (man)
 C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
 C:Accession: I38756
 R:Lee, R.A.; Marfaglia, S.M.; Branton, D.; Chishti, A.H.
 Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994

QY 475 ITS--RDVTIGSAPLYVKNILPRGAIIODGRKAGDRIEYNGVDLVKSGQEEVYSLR 532
 Db 217 IAGVGQOHIPGDNSTIVTKIIEGGAHMDGRLOIGDKLAVNSVGLDVAHVEDVAALK 276
 QY 533 STKMEGVSLVFERQ-----EDAFHREINAPESQOMIPEKTK-----AEDEDVLPPDG 562
 Db 277 NT--YDVVYLKAVKPSNATLSDSVAPDITTSYSQ--HIDNEISHSVGLGDYPIAMTPYS 333
 QY 583 TREELTF-----EVP-----LNDGSGAGLVGVSNGSKNNHADLGIPEKSIING 628
 Db 334 PRRTSVANOLLGEEDIDPREPRIYIHRSGLGPNVIGEDGE-----GIFSLIAGG 388
 QY 629 AASKDGRLEVNLDLAVNGESLLGKTNDAMETIRSMSTEGKRCMIOLIVARRISKON 688
 Db 389 PADLSELGRKQDQILISVNGVDLRNASHQAALAKNA-----GQVYIIAQ----- 434
 QY 689 ELKSGSPPELPETLALDD--RERRISHSLYSGIEGLDESPRNALSRINGESGKYQ- 746
 Db 435 -----YKPEYSRFEAKIHDLRQLMNSSLGSGTASLNSPRGFTY--RALFDYDKTK 487
 QY 747 ---LSPYVNMFPQDVTYIIEEDRLVLPRLHSDQS-----SSSHDDVGVFTR-----D 791
 Db 488 CGFSQALSPFEGVLAVID-----ASDEEMWQARVHSDSETDIDGIFPSKRKRYE 538
 QY 792 AGTAKKAAISDSADCSLSPD--VDPVLAFO 819
 Db 539 RREMSRLKAKDWGSSSGQREDSVLSTE 567
 RESULT 15
 T43275
 N:Alternat names: actin-binding protein
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text change 11-May-2000
 R:Nakaniishi, H.; Ohsaishi, H.; Satoh, A.; Wada, M.; Mandai, K.; Sato, K.; Nishioke, H.;
 J. Cell Biol. 139, 951-961, 1997
 A:Title: Neurabin: a novel neutral tissue-specific actin filament-binding protein involve
 A:Reference number: 22384; MUID:98011926
 A:Accession: T43275
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1095 <NAK>
 A:Cross-references: EMBL:U72994; NID:g2623756; PID:g2623757; PIDN:AAC53454.1
 A:Experimental source: brain

Query Match 4.2% Score 293; DB 2; Length 1095;
 best Local Similarity 19.5%; Pred. No. 6.1e-07;
 Matches 269; Conservative 163; Mismatches 477; Indels 470; Gaps 56;

QY 75 LVAVPEODPHHGCGTASSTGTSPETFGSELGTN---NY---SAPQYQATSEIE 126
 Db 30 LKSTPKPRP---DGEOKTEGEGSQOSRGRTKYSVNTKILFMQMGMEPNENAAIIA 85
 QY 127 VV-----PSVLRANM-PLHVRSSDPALIGLSTVSQD--SNFS--EES--RKN 169
 Db 86 KTRGGRSSPOKRRKPREVEKTDGVSVKLESSVERISRFDTMDGVSYAKTFETKRA 145
 QY 170 PTRMSTGTGLKONTAGSPKTCRRKDKENYRLPRD-----TSMNSNQQRMASSISA 224
 Db 146 FERSHREG---QNNRHS-----KKEKAGEAPQDEMGGSKNSGSSLSLSPRTQA 197
 QY 225 SHPWYKMLEKQEDDED-----GTEEDNSRVEPVGHADTGLEHIFNFSLIDMWKLVYEPN 279
 Db 198 VSPVSOQLSAVFENSESGAITPGKAKNSNYVTGHTPLNPSVTVMID----- 247
 QY 280 DGGFLGHVVPFARGKRTLGILVKRLKRGKAHEHNLFPRENDCLVRIINDDLANKRPFQ 339
 Db 248 -----TFG-----HLKDSNSRPSNNKQ 264

QY 340 A-----QHMFQAMRPTLIWEHVVPANKKEQYEQDLSQSEKNNYSSRFPDSQYIDNRGVN 395
 Db 265 ATDFEEPRKSAVYVPEVAOKGTSLASLPSEERQLSTEAEDVTAQPDVPTDSDKQSGEP 324
 QY 396 SAGLHTYQAPRLNHPPEQID-----SHSLPLPSAIPSG-----KPPSAPASA----- 438
 Db 325 SAESQAMPKSNITLSRPEPLEDAEAVVSGSEADQFORRDLDGGDLPSPDASASSCGKEV 384
 QY 439 --PQNVFTT-----VSSGVNKKIKGRNLQL--KKGTGELGFSITSRDVTIGGSAPIYKN 492
 Db 385 PEDSNSEGSVHYMHSDYVAVYVRSRYSNDMEGCTE----- 421
 QY 493 ILPRGAIIQDGRKAGDRIEYNGV-----DLVKSQEEVYSLASTMEGTVSLVF 545
 Db 422 -----QD-----EGDSDDENNYVQPMEXSELVGLPQEBELPANKKTF--SCAPIKVF 468
 QY 546 R--QEDAFHREINAPESQOMIPEKTKAEDDVLVTPDGTREFLTFEVLNDGSA----- 599
 Db 469 NTYSNEDYDRNDVDVPAASAEYELERKVELLFP-----VELEKDEGLGIST 520
 QY 600 GLGVSVKNGRKENHADLGIPEKSIINGCAASKDGRLEAVNDOLIAVNGESLLGKTNDAM 659
 Db 521 GMGVADAGLEK-----LGIFVKTYEGGAORDGHAIQVNDQIVEVDGISTLVGTQNPAA 575
 QY 660 ETLRSMSTEGNKRGMIOIVARRISKCNELSPGSPGPPELPETLALDD--RERRISHS 717
 Db 576 TVLR-----NRKNVRYIGR-----EKPGVSEVAQLISQTLQERRORELLER 620
 QY 718 LYSIGIEGLDESPRNALSRINGESGKYQLSPTVNMFPQDVTYIIEEDRLVLPRLHSDQS 777
 Db 621 HYAQYADDD-----ETGEYATDDE-----EDEGPTLP----- 649
 QY 778 SSSSHDDVGVTAADACTWAKAA--ISDSADCSLSPDVVLAFOREGFGQMSSEKTKQ 835
 Db 650 -----GGDAIEVFELEPNEDSPSDLD-----TSKLSHR 680
 QY 836 FSDASQLDVFKTRKSKMDLGIADETKLTNTVDQKAGSPRDVPSLGUKSSLESIDT 895
 Db 681 FKEL-QIKIAVTE-----AETOKLKT-----KIDA 704
 QY 896 AVAEYTLMGDIPFHRPRRIIRGRGNESFRAAIDKSYKPAVDDDECMETLEEDTERS 955
 Db 705 AE-----NEKYRWELKKQLOQNTENKEMVKL----- 733
 QY 956 SRSGRESVSTASQPSHSLERQMNQNEKGDKTDRKKDTGK----- 997
 Db 734 -----ESTWIEAQTLCHTVNEHLKETQSOYQALEKKYNKAKKLINDPOKELDPIRROEV 788
 QY 998 EKKRDROKEDKMKAKKMLKLGDM---FR-----GKHKKD 1032
 Db 789 EKKLLEVEREKAHLVEVOGLQVIRIDLEAVFPALLKQNTQVNNNNNIFERRPSGEVSKG 848
 QY 1033 DKTE--KTGRIKIQESFTS-----EEERIMKQEOBRIGIAKREFERRQARPDYAEI 1083
 Db 849 DTMEENEVAKQTSQDGLSDLINEAVPETERLDSKALKTRALQSVKNRQAPTRIR----- 903
 QY 1084 QDPHRTFGDDDELMTGCVSSYSGSMALNAR-----PQSPREGMMALYLAQYKKPR 1134
 Db 904 -----LVDVSVSTDGEDSLERKNFTFNDDPSFSSSSAD--LSGLAEKPKTG 949
 QY 1135 NKRPFSEVDSNRSTPSNHRIOQLRQFOQAKODVEDVDRRRYTSFQOPPMNARPAPOQSR 1194
 Db 950 LSQSLALSSDESL---DMT-----DDEITLD-----GQSPK 977
 QY 1195 HSAVVEYQMOQROEBRSSQAOAROYSSLPQSRKNASSVSQDSMEQ-----NYSQK 1248
 Db 978 HTQS-----OSRAVHEMSVOOVSHWLVGLSLDQYVSEPSAONISGEOLLOIDGKKLAL 1031
 QY 1249 GFQAKENPRYTSYQSGSRNGYTLGCHGNARVHLEFQELLROQRKKEQO--MKROPSE 1305
 Db 1032 GMTSSODRALVKKIKLKMKSLE---KARKQOEKQERKRLRKEQEBQMQRSKKSE 1086

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 14:29:54 ; Search time 24.98 Seconds

(without alignments)
2101.830 Million cell updates/sec

Title: US-09-757-781-2

Perfect score: 7035
Sequence: 1 MKVTVCFGRTVVVPCGDGH.....SOVARLRNLQTPKGRPFYS 1356

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	324.5	4.6	2485	1	PTND_HUMAN
2	296	4.2	1816	1	AF6_HUMAN
3	294	4.2	849	1	DIG3_MOUSE
4	293	4.2	767	1	DIG4_HUMAN
5	293	4.2	849	1	DIG3_RAT
6	293	4.2	1095	1	NEB1_RAT
7	291	4.1	724	1	DIG4_MOUSE
8	286.5	4.1	904	1	DIG1_HUMAN
9	286	4.1	724	1	DIG4_RAT
10	286	4.1	870	1	DIG2_HUMAN
11	284.5	4.0	817	1	DIG3_HUMAN
12	279	4.0	852	1	DIG2_RAT
13	278.5	4.0	911	1	DIG1_RAT
14	246	3.5	2492	1	ATRX_HUMAN
15	238.5	3.4	742	1	NEB1_HUMAN
16	238.5	3.4	2805	1	MAPA_HUMAN
17	237.5	3.4	3924	1	ANK2_HUMAN
18	233	3.3	1411	1	TGOF_HUMAN
19	229	3.3	1462	1	NKCR_HUMAN
20	228	3.2	2464	1	MAPB_MOUSE
21	227.5	3.2	960	1	DIG1_DROME
22	221.5	3.1	771	1	CAUD_CHICK
23	221.5	3.1	2774	1	MAPA_RAT
24	217	3.1	2426	1	SON_HUMAN
25	214.5	3.0	1190	1	ZO2_HUMAN
26	213	3.0	1167	1	ZO2_MOUSE
27	212	3.0	1004	1	MOT_HUMAN
28	210.5	3.0	1174	1	ZO2_CANFA
29	210.5	3.0	2459	1	MAPB_RAT
30	209.5	3.0	1189	1	YJH6_YEAST
31	209.5	3.0	1898	1	TRHY_HUMAN
32	208.5	3.0	817	1	NEB2_RAT
33	206	2.9	2845	1	APC_MOUSE

34	204	2.9	1021	1	MAPA_MOUSE	09qy6 mus musculus
35	203.5	2.9	1233	1	YF16_YEAST	P43597 saccharomyc
36	203.5	2.9	2843	1	APC_HUMAN	P25054 homo sapien
37	203	2.9	1453	1	NKCR_MOUSE	P30415 mus musculus
38	203	2.9	1654	1	PCPB_HUMAN	094913 homo sapien
39	203	2.9	1781	1	AKAC_HUMAN	002952 homo sapien
40	202	2.9	3969	1	HRX_HUMAN	003164 homo sapien
41	201	2.9	2476	1	ATRX_MOUSE	061687 mus musculus
42	199.5	2.8	1023	1	GLT_DROME	P33438 drosophila
43	199.5	2.8	1531	1	NPM5_HUMAN	094916 homo sapien
44	197.5	2.8	1744	1	TANA_XENLA	001550 xenopus lae
45	196.5	2.8	1359	1	ATRX_CAEEL	09u7e0 caenorhabdi

ALIGNMENTS

RESULT 1
PTND_HUMAN STANDARD; PRT; 2485 AA.
AC Q12923; Q15263; Q16826; Q15264; Q15265; Q15159;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Protein tyrosine phosphatase, non-receptor type 13 (EC 3.1.3.48)
DE (Protein-tyrosine phosphatase 1E) (PTP-EL) (PTP-BAS) (Protein-tyrosine phosphatase 1)
DE (FAP-1).
GN PTPN13 OR PTP1E OR PTP1L OR PNP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=94350988; PubMed=8071359;
RA Banville D., Ahmad S., Stocco R., Shen S.-H.;
RT "A novel protein-tyrosine phosphatase with homology to both the
RT cytoskeletal proteins of the band 4.1 family and junction-associated
RT guanylate kinases.";
RT J. Biol. Chem. 269:22320-22327(1994).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Leukemia;
RX MEDLINE=94116679; PubMed=8287977;
RA Maekawa K., Imagawa N., Nagamatsu M., Harada S.;
RT "Molecular cloning of a novel protein-tyrosine phosphatase containing
RT a membrane-binding domain and GluF repeats.";
RT FEBS Lett. 337:200-206(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=95014139; PubMed=7929060;
RA Sars J., Claesson-Welsh L., Heldin C.-H., Guez L.J.;
RT "Cloning and characterization of PTP1L, a protein tyrosine phosphatase
RT with similarities to cytoskeletal-associated proteins.";
RT J. Biol. Chem. 269:24082-24089(1994).
RN [4]
RP SEQUENCE OF 1216-2490 FROM N.A.
RC TISSUE=Pancreas;
RX Wang H.Y.;
RN Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases.
RN [5]
RP STRUCTURE BY NMR OF 1361-1456.
RX MEDLINE=20170882; PubMed=10704206;
RA Kozlov G., Gehring K., Ekfel T.;
RT "Solution structure of the PDZ domain from human phosphatase hPTP1E
RT and its interactions with C-terminal peptides from the Fas
RT receptor.";
RT Biochemistry 39:2572-2580(2000).
CC -I- FUNCTION: BINDS TO A NEGATIVE REGULATORY DOMAIN IN FAS THAT
CC INHIBITS FAS-INDUCED APOPTOSIS.

CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES WITH THE EXCEPTION OF
 CC THE LIVER AND SKELETAL MUSCLE. MOST ABUNDANT IN LUNG, KIDNEY AND
 CC FETAL BRAIN.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12128; AAB60339.1; -
 DR EMBL; D21209; BAA04750.1; -
 DR EMBL; D21210; BAA04751.1; -
 DR EMBL; D21211; BAA04752.1; -
 DR EMBL; X80289; CAA56563.1; -
 DR EMBL; X79676; CAA56124.1; -
 DR PDB; 3PDZ; 17-MAR-00.
 DR MIM; 600267; -
 DR InterPro: IPR000299; Band_4.1.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR000387; Tyr_phosphatase.
 DR InterPro: IPR000242; Tyr_prot_phosphatase.
 DR Pfam; PF00373; Band_4.1; 1.
 DR Pfam; PF00595; PDZ; 5.
 DR PRINTS; PR00935; BAND4.
 DR PRINTS; PR00700; PRTPHPTASE.
 DR SMART; SM00295; B4; 1.
 DR SMART; SM00288; PDZ; 5.
 DR SMART; SM00194; PTPc; 1.
 DR PROSITE; PS00660; BAND_4.1; FALSE_NEG.
 DR PROSITE; PS00661; BAND_4.1.2; FALSE_NEG.
 DR PROSITE; PS50057; BAND_4.1.3; 1.
 DR PROSITE; PS50106; PDZ; 5.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; FALSE_NEG.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE; 1.
 KW Structural protein; Cytoskeleton; Hydrolase; Repeat; 3d-structure;
 KW Alternative splicing; Coiled coil.
 FT DOMAIN 56 59
 FT DOMAIN 585 879
 FT DOMAIN 2237 2485
 FT DOMAIN 379 399
 FT DOMAIN 469 504
 FT DOMAIN 1775 1804
 FT DOMAIN 2057 2085
 FT DOMAIN 1093 1178
 FT DOMAIN 1368 1452
 FT DOMAIN 1501 1588
 FT DOMAIN 1788 1868
 FT DOMAIN 1882 1965
 FT DOMAIN 1742 1749
 FT ACT_SITE 2408 2408
 FT VARSPIC 884 1074
 FT VARSPIC 1056 1074
 FT CONFLICT 1134 1135
 FT CONFLICT 1216 1229
 FT CONFLICT 1238 1239
 FT CONFLICT 1357 1357
 FT CONFLICT 1362 1363
 FT CONFLICT 1383 1383

FT CONFLICT 1538 1538 R -> A (IN REF. 3).
 FT CONFLICT 1649 1649 R -> K (IN REF. 4).
 FT CONFLICT 1698 1714 KSDPTCTMYYDQKI -> RVKIPVPCFTLRK
 FT CONFLICT 1797 1797 (IN REF. 4).
 FT CONFLICT 1856 1857 G -> A (IN REF. 3).
 FT CONFLICT 2069 2069 AA -> G (IN REF. 4).
 FT CONFLICT 2206 2210 A -> S (IN REF. 4).
 FT CONFLICT 2485 AA; 276903 MW; 8DIB31597C6962B CMC64;
 SQ SEQUENCE

Query Match Best Local Similarity 4.68; Score 324.5; DB 1; Length 2485;
 Matches 193; Conservative 134; Mismatches 361; Indels 245; Gaps 32;

QY 80 DEQDPHGGDGNSSASTGTOSPELFGSLG-----TNNVAFQPYQATSEI--E 126
 DB DAQDI-----EASPRSLNLQAEVYGFNMGRALSTGSLASTINKLAVRPLVQAEILKR 934
 QY 127 VTPSVLRANMPLH--VRRSDPAL-----IGLSTVSQDSNFSESEPSRKPTKMTAG 179
 DB 935 LSCSELSTYPLQNSRKKEKNDKASWEKPRMKSYSYHDSQALYHRKN----- 984
 QY 180 LKONTAGSPKTCDDKKDENTRSILPR-DTSWMSQFORDNARSSLASHPWGKMLEKQEQ 238
 DB 985 VIVNMEPPQVVAELVCKPQHMSRSDAESTLAGVTYRLNNSKVASLIN-----RSP 1034
 QY 239 DEDCTEDNSRVEEVGHA-----DTGLEHIFNFSLDQV-----VALIV 275
 DB 1035 ERRKHESSSIDEDPGQAYVLGNTMHSQSSSQVYLKENDYLHKRWSTVSPEREITLY 1094
 QY 276 EVPDGS-PLGIHVPPSANGRTLLYLKLEGGKAEHMLFENDCIYRINDGILRN 334
 DB 1095 NLKKDAATGIGFQITIGEEKMGRILGIFISVAPGAPADLDCILKPGRLISVNSVSLG 1154
 QY 335 RREFQAHMFRQA-----MRTPILMFHV-----VPAA 361
 DB 1155 VSHHAILEILQNAPEVDYTLVISOPEKIKVSTVYHLNENKNYKRSYMODAIDSS 1214
 QY 362 NKROY-----EQLSSEKNYYSRPF----- 382
 DB 1215 SKDHMSRQTLAHISENSFGPSGGLREGSLSSODRTESASLSQSYVNGFASHILGQTW 1274
 QY 383 -----SPDSQYIDNRSVYASAGLHTVQ---RAPRLNHPEPID-SHSLPLSAHSPKRP 432
 DB 1275 QRSQHSPPSPVYSKATEKEFTPDNOSKTKKPGISDVTYDSRGDSDMDENATYSSQDH 1334
 QY 433 SAPASAPQVVFSTYSSGYNT-----KIKGRNLIOLEKTEGIGFESIT---SPDVTIGG 484
 DB 1335 QTPKQESSSVNTSKRMKFTSSPPKPGDIFVEELAKNDNSLGISVGTGVMTVARHG 1394
 QY 485 SAPIVYKNILPRGCAIIDGRILKAGDRILEVGVPLVNGSOEYVSLSRKMGTVSLV 544
 DB 1395 ---IYKAVIPQGAESDGRHKKDRILAVANGVSLGATIKQAVETLRNT---GQVYVHL 1448
 QY 545 FROEDAFHRE-----LMAEPSQMOIP---KETRAEDDIYLPDGTREFLFEV 591
 DB 1449 LKQOSPTSKHEHVPDQCTLSDQNAOGGPEKVKTYQKXDSFV-----TEENTFEV 1502
 QY 592 PL-NDSSAGLGVSVKGRSKENHADLGIFVKSTINGGAASKGRILVNDQLAVNGSL 650
 DB 1503 KLFKNSGGLGFSRENNLLPEQINMSIYAVKRLPGQPAESGKIDVDVILKVGASL 1562
 QY 651 LKTNQAMETLRRSMSTEGNKRGMQLIVARRISCKNELKPSGSPPEGLPIETALDDR 710
 DB 1563 KGLSQQEVISALGTAPF-----VFLLCR-----PPGVLPETIDAL--- 1600
 QY 711 ERRISHLYGIGIGLDESPRNALSR-----MGESGYQL-SPTVNMPODDTV 759
 DB 1601 ----LPLQSPQAVYLPNSSKDSQPCVQOSTSSDENEMSKKCKCKSPSRSDSYSSS 1656
 QY 760 ILEDRLPYLPLPHLSDQS-SSSSHVDVGFVTADAGTWAKAIS----- 801

DB 1657 GSGEDLTATAPANINSNWSSALHOTLSNMVSOAQSHHEAPKSOEDTICTMYFPQKIPN 1716

QY 802 -----DSADCSLSPDVDPYLAFOREGFGROSMS 829

DB 1717 KPEFDSNPSPLPPDMAPOQSQYPOSESASSSS 1749

RESULT 2

AF6_HUMAN STANDARD; PRT; 1816 AA.

AC P55196; 075087; 075088; 075089; Q9NU92;

DT 01-OCT-1996 (Rel. 34, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE AF-6 protein.

GN MLLT4 OR AF6.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RX MEDLINE=94061833; PubMed=8242616;

RA Pressed R., Gu Y., Alder H., Nakamura T., Canaan O., Saito H., Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y., Croce C.M., Canaan E.;

RA "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in acute myeloid leukemias with the t(6;11) chromosome translocation.";

RL Cancer Res. 53:5624-5628(1993).

RN [2]

RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.

RC TISSUE=Fetal brain;

RX MEDLINE=98344142; PubMed=9679199;

RA Saito S., Matsushima M., Shitahama S., Minaguchi T., Kanamori Y., Minami M., Nakamura Y.;

RA *Complete genomic structure, DNA polymorphisms, and alternative splicing of the human AF-6 gene.*;

RL DNA Res. 5:115-120(1998).

RN [3]

RP SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).

RA Williams S.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT CONTROLLED BY RAS SIGNALING PATHWAYS.

CC -1- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCUDIN.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1, 2 (SHOWN HERE) AND 3; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL TRANSLOCATION t(6;11)(Q27;Q23) THAT INVOLVES MLLT4 AND MLL/HRX.

CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.

CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch).

CC -----

CC EMBL; AB011399; BAA32484.1; -

DR EMBL; AB011399; BAA32484.1; -

DR EMBL; AB011399; BAA32485.1; -

DR EMBL; U02478; AAC50059.1; -

DR EMBL; AL049698; CAB/6850.1; -

DR HSSP; Q12923; 3PDZ.

DR MIM; 159559; -

DR InterPro: IPR002710; DIL.

DR InterPro: IPR000253; FHA_domain.

DR InterPro: IPR001478; PDZ.

DR InterPro: IPR00159; RA.

DR Pfam: PF01843; DIL; 1.

DR Pfam: PF00498; FHA; 1.

DR Pfam: PF00595; PDZ; 1.

DR Pfam: PF00788; RA; 2.

DR ProDom: PD003376; DIL; 1.

DR SMART; SM00240; FHA; 1.

DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS0106; PDZ; 1.

KW Chromosomal translocation; Proto-oncogene; Alternative splicing.

FT DOMAIN 36 206

FT DOMAIN 425 491

FT DOMAIN 804 910

FT DOMAIN 991 1077

FT DOMAIN 162 174

FT DOMAIN 1349 1356

FT DOMAIN 1371 1376

FT DOMAIN 1561 1571

FT SITE 26 26

FT VARSPLIC 1588 1611

FT VARSPLIC 1612 1816

FT VARSPLIC 1666 1743

FT VARSPLIC 1744 1816

FT CONFLICT 373 373

FT CONFLICT 391 391

FT CONFLICT 744 744

FT CONFLICT 1031 1031

FT CONFLICT 1408 1408

SO SEQUENCE 1816 AA; 205604 MW; EB1FE7F04879CEBF CRC64;

Query Match 4.2%; Score 296; DB 1; Length 1816;

Best local similarity 20.1%; Pred. No. 9.8e-07;

Matches 340; Conservative 217; Mismatches 536; Indels 600; Gaps 86;

QY 49 HLEHGDGGLDLDLDCV-----ADKRLVAVFDEDDPH----- 86

DB 307 HSDKAKAKERILDDDECPQLIFREWPSDKGLVFLKRPDPHFKTKKHLEGKTPKG 365

QY 87 -GGDGSAST-----GTQS---PEIFG-----SELGT-----NNVSAF 116

DB 366 KERADSGYGSTLPPKRLPYVELSPDGSQSDKPKLYRLQLSVTEGTEKLDNSIQLE 425

QY 117 ---QPYQA---TSEIEVTPSVLRANMPLHYRSSDPALI--GLTSVSDSN-FSSSE 164

DB 426 GPGLHHODLTMMGCVVYTPRSMDAEPYVEQRISETTMMQSKNVQGCASHVKEFVD 485

QY 165 PSKKNP-TKMSSTTAGFLKONTAGSP-----KTCDRKKD-ENTRSLPRTSMMNSQFORDN 217

DB 486 PSDNDHLARKSVYDGLMVGPRHKPGIVQETTPDLGDIHSGTALP--TSKSTTRLSDR 543

QY 218 ARSLASASH-----PMVG-----KMLEKQEDDEGTE-----EDNSRVEP 252

DB 544 VSSASTAEKGMVKPMIRVEQOPDYRQESRTDASGPELILPASIEPRESSDSLSAI 603

QY 253 VGHADTGLEHI---PNFSLDMMVKLEVPRDGPGLGIHVFPFSGRGRTGLLVKRLKNG 309

DB 604 INTNSGVHFKLSPIYVLMACRYV-LSQYRP---DISP-TERHNKVLAVVKNKV--- 655

QY 310 GKAEHNLRENDCLV-----RINDGDLNRKFEQAQHFROAMKPTIIVHVPAAK 363

DB 656 --SMMEGVIOKKNIGALAFMANNASELIN-FIKODRDLSTRITLDAQDLAHLVOMARK 712

QY 364 EGYEOLSSSEKKNYSSRSPSDQYIDNRVNSAG-----LHTQO-----RARLNH 410

DB 713 YLVHCL-QSELNMYMPA-----FLDDPEENSLQRPKIDVDLHTLGLAMSLILRCRVN- 763

OY	411	PEBODISRLPHSAPHSGKPSAPASAPAWFSGTWS	---SGYTRKTKGRKLNOL---	464	
Db	764	AMTLTOLSQLFPHN	-----MFLFNLTVPDPSGLCSHWGALITROOLHIT	810	
OY	465	---KKCEBGEGESTSRDY	-----TIGGSAPITVYNI	870	
Db	811	EAWEKGELELAADCHSRIVQATFTLLTMKRAPDDININSTCFKINSLOALALONH	-----LPRGAATODGR	504	
OY	505	LKAGDRLIEYNGVDLYGKSOEEVYSLRSTKMGVSLVLPQODAPHPREIAPNSQMO	-----	564	
Db	871	CAPEPEIP	---TDLI---ENVYVAENADE---	908	
OY	565	IPKFKKEDBDIVL	-----TPDGTREBEFL	588	
Db	909	---EDPDLOQLPFLLPEDGSCDVRNIPNGLQEFLLPDLQCGFCRLPHRSGTW	-----	961	
OY	589	---FE---	-----VPLNCSGAGIG-VSVKGRSKENHNDL	617	
Db	962	TIYFEBADYESHLRENTLEAOLPKEPELITVTLKONGMGLSTIAAGK	-----AGQDKL	1017	
OY	618	GTFVSEITINGSAASDGRLEBYNDOLIANGESLICKTODMELTRKSMST	-----EGNRKG	674	
Db	1018	GTYKSVYVGGAAD-DGRLAGDOLLSDVSRSLVNGLSQERAEALMTRTSSVTLLEVAKOG	-----	1076	
OY	675	---MIOLIVARRISKCNELASP	-----GSPPEL	701	
Db	1077	AIYHGLATLINOPSYMMORI	SDRRSGKPRKREGEFLYNSTONGSPESPOLPMAETSE	1136	
OY	702	PIETALDR-ERRISHLYSGIEGLDPSRMAALSRIMESCKOYLSPIVNM	-----	753	
Db	1137	PKRLGDDRLMKRNADHRSPVNAOPSPGKAYA	---SGTAKITTSYGNCTEBOT	1194	
OY	754	---PODDTYI---	-----EDDRLEVP	784	
Db	1195	PPRPBAYIPIQYTRFTFTFPASSQDR	---APPONMENEKPRHMTDINSMS---	1249	
OY	785	VGEVTAADACTWAKAALISDADCSLSPVDVPLAPRQEGROSSEKRTKOPSDASQDF	-----	844	
Db	1250	---	---IACQVTRSGEALREDKAYOL-ERRITA	1277	
OY	845	VKTRKKSAMDIGLADETKL	---NTYDDKAGSPRDVGPGLGKKS---	SSLESIO	894
Db	1278	AMDRSOS-DMMINOSSLSDSTSOEHLNHSISVTPASTLTKGPGRWKTPAIPATP	-----	1336	
OY	895	TVAEVLINGDIPHRPRRILLRGCSNEFRAIDKSYKPAVNDDBEGMETLEEDTEE	-----	954	
Db	1337	VAVSO-PIRDLTPPPPPV	-----HYAGDDGMSMDLPLPPPS	1376	
OY	955	SSRSRESYASDOPSHLEROMONOKGKTDRRKDKTGKKKODKCKOKMAKK	-----	1014	
Db	1377	ANQIGLPSAOVA	---ARRRREHOMWEKEKARLEE-	1412	
OY	1015	GMKGLGDMFRGKHKRDKITEKGIKO	-----ESFSEBERIKMOBERIAKT	1067	
Db	1413	RE-FREROAREBDAEIODFHRTGDCDELMAGVSYSEGSMAINAP	-----	1114	
OY	1068	RE-FTREROAREBDAEIODFHRTGDCDELMAGVSYSEGSMAINAP	-----	1157	
Db	1457	CEYVIRLELOPOQR	---RTERRD-LQYITYSKELSSGDSLSPDPWKADAKEL	1507	
OY	1115	OSPREGMMDALYAOYKPRNKSPPVDNSRTPSNHDIORLOPFO	-----	1162	
Db	1508	EKOQOMHAYDMLSKEIFE-LOSKP	---DRAESB-DRLKRLMLEMOCKRLDESOKD	1560	
OY	1163	---QAKODEV---	---EDRRRTYSFQPPMNPAPNPTOGGRHSVSE	1200	
Db	1561	EDDEEEDDDVYTMILMORLEARRARLODEKRRQOOLE	---MKRREADARQ---	1613	
OY	1201	VOMORORERSSOQARQ	---YSLPROSRK---	NASSYQSDSWE	1241
Db	1614	EEFRRRROEERTKDAEKKRROEGYSLTEFRRRRODEARLLEPAPGLCPPLP	-----	1672	

RESULT	3		PRT:	849 AA.
ID	DIG3_MOUSE	STANDARD:		
AC	P70175;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Presynaptic protein SAP102 (Synapse-associated protein 102) (Discs,			
DE	large homolog 3).			
DE	DIG3 OR DIGL3.			
GN	Mus musculus (Mouse).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus-			
OX	NCBITaxID=10090;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6; TISSUE=Brain;			
RA	Kohmura N., Makino S., Yagi T.,			
RL	Submitted (AUG-1996) TO THE EMBL/GENBANK/DDJB DATABASES.			
CC	-1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR SUBUNIT NR2B. CONTAINS 3 PDZ/DHR DOMAINS.			

DR	EMBL; D01016; IBB9.	
DR	HSSP; P21016; IBB9.	
DR	MGD; MG11089986; D1gh3.	
DR	InterPro; IPR000619; Guanylate_kin.	
DR	InterPro; IPR001478; PDZ.	
DR	InterPro; IPR001452; SH3.	
DR	Pfam; PF006635; Guanylate_kin; 1.	
DR	Pfam; PF005955; PDZ; 3.	
DR	Pfam; PF000148; SH3; 1.	
DR	SMART; SM00072; GKXC; 1.	
DR	SMART; SM00228; PDZ; 3.	
DR	SMART; SM00326; SH3; 1	
DR	PROSITE; PS00856; GUANYLATE_KINASE_1; 1.	
DR	PROSITE; PS50052; GUANYLATE_KINASE_2; 1.	
DR	PROSITE; PS50106; PDZ; 3	
DR	PROSITE; PS50002; SH3; 1.	
KW	SH3 domain; Repeat.	235
FT	DOMAIN	149
FT	DOMAIN	244
FT	DOMAIN	330
FT	DOMAIN	404
FT	DOMAIN	519
FT	DOMAIN	589
FT	DOMAIN	659
FT	DOMAIN	849
SEQUENCE	849 AA; 93482 MW; EF3EF2D7513538EE	CRC64

score 294; DB 1; Length 849

FT	DOMAIN	203	289	PDZ 2.
FT	DOMAIN	356	436	PDZ 3.
FT	DOMAIN	471	541	SH3.
FT	DOMAIN	577	767	GUANYLATE KINASE.
FT	CONFLICT	46	46	E-> V (IN REF. 3).
FT	CONFLICT	81	83	V-> EFR (IN REF. 4).
FT	CONFLICT	399	401	GDO -> AGI (IN REF. 4).
FT	CONFLICT	767	85429	MM; BE10191595652D8 CRC64;
SO	SEQUENCE			

4.2% Score 293; DB 1; Length 767;

Query Match Best Local Similarity 21.8%; Pred. No. 4.4e-07; Mismatches 150; Conservative 150; Indels 233; Gaps 28;

QY	210	SNOPR	-----DNARSSLSASHPWGKLEKODODEGTEEDNSRYEPVGHADTGLEHIP	264
DB	32	SDLFQALDILDYVNASISESQ	-----KRYODED-----TPLEHSPA--HLP	73
QY	265	NFS	-----LDMVKLVEVPNDGPGGLIHVPPSARGRT	298
DB	74	NOANSPVIVNTDLEAPCYELQVNGTEGEMEYEETLERNSGL--GFSIAGIDNPH	130	
QY	299	LG	-----LLVRLERKGAHEHNLFPENDCIVRINDGDLRNRFEQAQHMFRQARTPIW	354
DB	131	IGDPSIFIKIIPGMAADGRLRYNDSILFVNEVDVREYHSAVALEKEA	183	
QY	355	PHVVPANKQEQYEQLSQSEKNYSSRFSPDSQYIDNKSNSAGLHTVORAPRLNHPPEQ	414	
DB	184	-----GSIVRLVWR	194	
QY	415	IDSHSRLEPHSAHPGKPPASAPAPQNVFTVSSGYNTKIKIGRLNIOLKGTGSGES	474	
DB	195	-----KPPAE-----KVAEIKILGPGGLGHS	216	
QY	475	INS	-----RDVTIGSAPIVYKNTLPKCAIIOGRKAKGRLELVNGVDLVGKSOEYVSLR	532
DB	217	TAGVGNQHIIPGDSIYTKIEGGAHAKRDLIGKILVNSVGLGDEVDMAVALK	276	
QY	533	STMEGVTLVLR	-----EDAFHRLNAEPPSOMOIPEYTK-----ADDEDIVLTPDG	582
DB	277	NT	-----YDVVYLKAKPSNATLSDSVAPDITTSYSQ-HIDNEISHSSYLGTGYPLAMTPS	333
QY	583	TREFLTF	-----EVP-----LNDSSAGLVSVYKNSRKNHMDLIFVKSITNG	628
DB	334	PRRSPYAKDLGEBEDIPREPRVIVHRSSTLGFINVIGDEGE	388	
QY	629	AASKDGRLVNDOLIAVNGESLIGKTNDAMETLRSMSTBCKRKMQLIVARISKCN	688	
DB	389	PADISGLKRGDOLISVNGVDLNMASHQALAKNA	434	
QY	689	ELKSGSPGPELPIETALDD-RRRISHLSYSGIEGLDPSNANLSTRIMESGYO-	746	
DB	435	YRDEKSRFEAKIIDLREOLMNSISGSTASLRNPKGYI-RALPYDXTKD	487	
QY	747	-----LSEPTNMPDODTVIIEEDRLPVLPHLSDOS	791	
DB	488	CGLTSLQALSFRRGDLVLIYD	538	
QY	792	ACTMAKAISDACSLSPD-VDPVLAFO	819	
DB	539	RREMSRLKAKMGSSSGSGGREDVLSYE	567	
RESULT 5				
ID	DLG3_RAT	STANDARD;	PRT;	849 AA.
AC	062936; P70547;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Presynaptic protein SAP102 (synapse-associated protein 102) (PSD-			
DE	95/SAP90 related protein 1) (Discs, large homolog 3).			
GN	DLG3.			

RESULT 5
DUG3_RAT STANDARD: PRT; 849 AA.
AC 062936; E70547; Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 01-NOV-1997 (Rel. 35, last annotation update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Presynaptic protein Sap102 (Synapse-associated protein 102) (PSD-95/SAP90 related protein 1) (Discs, large homolog 3).
GN DUG3.

OC	Rattus norvegicus (Rat).
OC	Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi;
OC	Mammalia: Eutheria: Rodentia: Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RN	(1)
RP	SEQUENCE FROM N.A. (LONG FORM).
RC	TISSUE=Brain;
RP	MEDLINE=96374358; Pubmed=6780649;
RX	Mueller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S.,
RA	Fenster S.D., Lau L.-F., Yen R.W., Hagan R.L., Gundelfinger E.D.,
RA	Garner C.C.;
RT	*Sap102, a novel postsynaptic protein that interacts with NMDA
RT	receptor complexes in vivo.;
RT	Neuron 17:255-265(1996).
RL	[2]
RN	SEQUENCE FROM N.A. (SHORT FORM).
RP	Itle M., Hata Y., Takai Y.;
RA	Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
RL	-1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC	SUBUNIT NR2B.
CC	-1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC	-1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC	-1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC	-1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@sib-sib.ch).
CC	EMBL; U50147; AAA93031.1; -
DR	EMBL; U53367; AAB48561.1; -
DR	HSSP; P31016; IBE3.
DR	InterPro; IPR000619; Guanylate_kin.
DR	InterPro; IPR001478; PDZ.
DR	InterPro; IPR001452; SH3.
DR	Pfam; PF00625; Guanylate_kin; 1.
DR	Pfam; PF00018; SH3; 1.
DR	Pfam; PF00018; SH3; 1.
DR	Pfam; PF00018; SH3; 1.
DR	SMART; SM00072; GUKG; 1.
DR	SMART; SM00328; SH3; 1.
DR	SMART; SM00328; SH3; 1.
DR	PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR	PROSITE; PS00856; GUANYLATE_KINASE_2; 1.
DR	PROSITE; PS50106; PDZ; 3.
DR	PROSITE; PS50106; SH3; 1.
DR	SH3 domain; Repeat; Alternative splicing.
KW	SH3 domain; Repeat; Alternative splicing.
FT	DOMAIN 149 235 PDZ 1.
FT	DOMAIN 244 330 PDZ 2.
FT	DOMAIN 404 484 PDZ 3.
FT	DOMAIN 519 589 SH3.
FT	DOMAIN 659 849 GUANYLATE KINASE.
FT	DOMAIN 627 640 MISSING (IN SHORT ISOFORM).
FT	VARSPLIC 627 640 MISSING (IN SHORT ISOFORM).
SO	SEQUENCE 849 AA; 93539 MW; 34DA9C46C7BB96DB CRC64;

4.2% Score 293; DB 1; Length 849;

Query Match Best Local Similarity 20.3%; Pred. No. 5.1e-07; Mismatches 184; Conservative 124; Indels 300; Gaps 35;

QY	264	PNFSL	-----DDMKLVEVPNDGPGGLIHVPPSARGRT	298
DB	134	PSLSVNGSDGMFYEEYELVLRGNSGL--GFSIAGIDNPHVDDPGIFIKIIPGGA	190	
QY	314	HENLFRENDICVIRINDGLRNRFEQAQHMFRQARTPIWPHVVPANKQEQYEQLSQSE	373	
DB	191	MDGRAGVNDCLRVNEVDV	209	
QY	374	KNNYSSRFSPDSQYIDNKSNSAGLHTVORAPRLNHPPEQIDSHSRLEPHSAHPGK	430	

```

Db 210 -----SEVYHSAVBALEAG-PVYRLVYRRQPP----- 238
OY 431 PPSAPASAPQNFSTVSSGYMTKKIGRLNIOLKKGTEGFSITS--RDVYIGSAPI 488
Db 239 -----PETI-----MEYNLKGPRGLGFSIANGIGNHIPPDSNI 273
OY 489 YKKNILPFGCAIQDGLKAGDRLEIYNGVDLVKSGOEYVSLRSTKMGVSLVFRGE 548
Db 274 YTKIIEGGAOKDRQLQIGDRLLAVNNNTNLDVREHEEVAASLKNT--SDWYLVKARP 330
OY 549 DAFHRELNAAERSQMQ-----IRK 567
Db 331 GSLHNDMAPPDYASTFTALADNRHISNLSGLYGAVESKVTYAPPOVPTRPSPIR 390
OY 568 ETKAEDEDIVLPDGTREFLTFEVP---LNDGSAGLGYGVKNGRKNENHADLGIPIVKSI 624
Db 391 IHLABE-----DPTRE-----PKIILHKGSTGLGFTNIGGEGE-----GIVVSTI 432
OY 625 INGAASKDGLRLVNDQILVANGESILGKTNDAMETLRRSMSTEGNKRQMIQILVARI 684
Db 433 IAGGPAIDLSGELRGDRILSTVNGVNLRNATHEQAAALAKRA-----GQSVTYVAQ-- 482
OY 685 SKCNELKSGSPGPELPLETALDO-REKRISHLSYSGIEGLDESPSRMAALSRIINGEG 743
Db 483 -----YRPEYSRPEESKIHDLREQMNMNSWSSGSLTSEKRSLLYVRALE---- 528
OY 744 KYQLSPVPMPODDTVIIEDDRPLPVLPHLSDQSSSSSSSDV-GEVYADAGTWAKAAI-- 800
Db 529 -----DYDRTRDSCLP-----SQGLSFSGDILHVNADDEWQAKRLV 568
OY 801 --SDSADCSLSPDVPVLAFOREGREGROSGMSERKTRKQFSDASQOLDEVKTRKSKMDLGI 858
Db 569 PHGESQIQLGIVP-----SKKRVEK-KERARKLVKFKHARTGMIESNR 609
OY 859 DETKLTNTVDQKAGSPSRDVGPSLGLKSSLSLEJOTAAEVLTLNDIDFHRPRPIIING 918
Db 610 DTPGLS--DDYGGAKNLKCVNTSNTSDSESS--KQGDALILSYEPVTRQBIHARPIILG 666
OY 919 ---RCGNESFRAIDKSY-----DKPAVDDDEGEM-----ETLEEDTEESS--RS 958
Db 667 PKMDVNDLILSEFPKHKFSCVPHTTRPRRDNEVDQDHFVYSRQMKDIDQNKFIET 726
OY 959 GR-----ESVSTASDQPSHSLEROMNGNOEKDKTRKKDKTGKKKDRDKED 1008
Db 727 GGFNDNLVGTSTIOSVRAVAERGHCI-LDVSGNAIK-----RLQQAOLYPI 771
OY 1009 KKKAKKMGKGLGDMFRFEGKHRRDKI-EKTKIKIQESETSEERIIRKQDERIQART 1067
Db 772 AIFIKPKSTIEALMKNROTYPQANKIFDKA--MKLEDFEGEFTYALVQGDLSLEIYKTI 829
OY 1068 REFERRQA 1075
Db 830 KOIIEQOS 837

```

```

RX MEDLINE-98031926; PubMed-9362513;
RA Nakanishi H., Obatahi H., Satoh A., Wada M., Mandai K., Satoh K.,
RA Nishio H., Matsura Y., Mizoguchi A., Takai Y.;
RT "Neurabin: a novel neural tissue-specific actin filament-binding
RT protein involved in neurite formation.";
RL J. Cell Biol. 139:951-961(1997).
RN
RN [12]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-Brain;
RX MEDLINE-98318661; PubMed-9653190;
RA Burnett P.E., Blackshaw S., Lai M.M., Qureshi I.A., Burnett A.F.,
RA Sabatini D.M., Snyder S.H.;
RT "Neurabin is a synaptic protein linking p70 S6 kinase and the neuronal
RT cytoskeleton.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8351-8356(1998).
RN [13]
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE-Brain;
RX MEDLINE-20054471; PubMed-10585469;
RA Macmillan L.B., Bass M.A., Cheng N., Howard E.F., Tamura M.,
RA Strack S., Madzinski B.E., Colbran R.J.;
RT "Brain actin-associated protein phosphatase 1 holoenzymes containing
RT spinophilin, neurabin, and selected catalytic subunit isoforms.";
RL J. Biol. Chem. 274:35845-35854(1999).
RN [14]
RN INTERACTION WITH TGN38.
RX MEDLINE-99445568; PubMed-10514494;
RA Stephens D.J., Bantling G.;
RT "Direct interaction of the trans-golgi network membrane protein,
RT TGN38, with the F-actin binding protein, neurabin.";
RL J. Biol. Chem. 274:30080-30086(1999).
RN [15]
RN INTERACTION WITH PPL, PHOSPHORYLATION SITE SER-431, AND MUTAGENESIS.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE-99435770; PubMed-10504266;
RA McAvoy T., Allen P.B., Obatahi H., Nakanishi H., Takai Y.,
RA Greengard P., Nairn A.C., Hemmings H.C. Jr.;
RT "Regulation of neurabin I interaction with protein phosphatase 1 by
RT phosphorylation.";
RL Biochemistry 38:12943-12949(1999).
CC -1- FUNCTION: BINDS TO ACTIN FILAMENTS (F-ACTIN) AND SHOWS CROSS-
CC LINKING ACTIVITY. BINDS ALONG THE SIDES OF THE F-ACTIN. MAY BE
CC INVOLVED IN NEURITE FORMATION. INHIBITS PROTEIN PHOSPHATASE 1-
CC ALPHA ACTIVITY. MAY PLAY AN IMPORTANT ROLE IN LINKING THE ACTIN
CC CYTOSKELETON TO THE PLASMA MEMBRANE AT THE SYNAPTIC JUNCTION.
CC -1- SUBUNIT: POSSIBLY EXISTS AS AN HOMODIMER, HOMOTRIMER OR AN
CC HOMOTETRAMER. INTERACTS WITH F-ACTIN, PROTEIN PHOSPHATASE 1 (PPL),
CC NEURABIN-II, TGN38 AND P70-S6K.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ALSO FOUND IN THE SOLUBLE
CC SYNAPTOTOMAL FRACTIONS.
CC -1- TISSUE SPECIFICITY: BRAIN, AND WIDELY EXPRESSED IN NEURAL TISSUE.
CC HIGHLY CONCENTRATED IN SYNAPSES OF DEVELOPED NEURONS. IN
CC DEVELOPING NEURONS, CONCENTRATED IN THE LAMELLIPODIA OF THE GROWTH
CC CONE.
CC -1- DOMAIN: INTERACTS WITH P70-S6K VIA ITS PDZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U72994; AAC53454.1; -.
CC HSSP: Q12923; 3PDZ.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001660; SAM.
CC Pfam: PF00595; PDZ; 1.
CC Pfam: PF00536; SAM; 1.
CC SMART: SM00228; PDZ; 1.
CC DR

```

DR SMART: SM00454; SAM; 1.
 DR PROSITE: PS0106; PDZ; 1.
 DR PROSITE: PS0105; SAM DOMAIN; 1.
 KW Neurogenesis; Actin-binding; Cytoskeleton; Coiled coil;
 KW Phosphorylation.
 FT DOMAIN 1 144 ACTIN-BINDING.
 FT DOMAIN 505 593 PDZ.
 FT DOMAIN 989 1052 SAM.
 FT DOMAIN 598 628 COILED COIL (POTENTIAL).
 FT DOMAIN 674 825 COILED COIL (POTENTIAL).
 FT DOMAIN 1036 1091 COILED COIL (POTENTIAL).
 FT DOMAIN 426 503 INTERACTS WITH PROTEIN PHOSPHATASE 1.
 FT DOMAIN 598 1091 INTERACTS WITH TGN38.
 FT MOD_RES 461 461 PHOSPHORYLATION (BY PKA).
 FT MUTAGEN 461 461 S->E: 35-FOLD DECREASE IN INHIBITION OF
 P1-ALPHA.
 FT MUTAGEN 516 517 GI->AA: ABOLISHES P70-S6K BINDING.
 FT SEQUENCE 1095 AA; 122735 MW; 58E3650320B3BD61 CRC64;

Query Match 4.2%; Score 293; DB 1; Length 1095;
 Best Local Similarity 19.5%; Pred. No. 7.1e-07;
 Matches 269; Conservative 163; Mismatches 477; Indels 470; Gaps 56;

QY 75 LVAVFEDDDHHGGDSTASSTGTGSPETFGSELTN-----NY-----SAFQPYATSEIE 126
 DB 30 LKSTPKPKP---DEOKTKEGEGSQSQRKKGNNRKLKLPQMOMGMPENNAITIA 85
 QY 127 VT-----PSVLRAAM-PLHVRSSDPALIGLSTVSQ--SNFSS--EPPS-----RKN 169
 DB 86 KTKGKGRPSSPQKRMKPKFEVEKTDGSGVYKLESSVSEKISFDMHGGPYAKFTFRKM 145
 QY 170 PTWMTTAAFLKONTAGSPKTCDDKRDENYRSLPRD-----TWNMSQFORARBSLSA 224
 DB 146 PERSHEG---QNNRHSB-----KKEKAGAEFPQDEMGSKSNRSSOSLGLSPRTBA 197
 QY 225 SHPNTGKMLEKQEDDED-----GTEEDNSRVEPYGHADTLEIIPNFSLDDMKLVPEYN 279
 DB 198 VSPVYSQLSAVFENSESGAITTPGKANSNYSVTGHTPLNLSVTVNTND----- 247
 QY 280 DGGPLGIHVYFSGRGRTGLLVKLEKGAHEHNLFRENDCIYRIDNDGLNRREPO 339
 DB 248 -----TFG-----RLKDNSRSPSSSKQ 264
 QY 340 A-----QMFQOAMPTPIHFHVYPAANKEQYBOLSOSEKNNYSRSPDSQYIDNRSVN 395
 DB 265 ATDTEPEKSEAVPEVAOKGSLASLPSEBROLSTAEIDVTAPDPDSTDKDSEPR 324
 QY 396 SAGLHTVQRAPLRNHPPEID-----SHSLRPHSAHPSC---KPPSAPASA--- 438
 DB 325 SAESQAMPKSNLTKRKEPLEDEANVVGSEAEQFORADLTGGDDLSPDASASSCKEY 384
 QY 439 --PQNFEST--VSSGYNTKTKIGRLNIQL-KKGTBGLSFTSIRVTTIGSAPLYKN 492
 DB 385 PEDSNFESHSYHMSDYNVFRSRYSDMGEGT----- 421
 QY 493 ILPRGATIDGRLKAGDLIEVNCV-----DIVKSOBEVVSLLRSTMEGTSLVLF 545
 DB 422 -----QD-----EGDSDSDNNYQPDMEYSELVGPQEEIIPARKIKF-SCAPIKVF 468
 QY 546 R--QEDAHFRELAEPSQOIPRETKAEDEDIVLPDGRRELTFEVLANDSGA--- 599
 DB 469 NYSNEDYDRNDVDVPAASAEVLEKREKLELP-----VELEKDDGISISII 520
 QY 600 GLGVSVKGNKSKENHADLGFVKSIIINGGAASDGRRLRVNDLIJAVNGESLLGKTNDAM 659
 DB 521 GMVGADAGLEK-----LGLFVATVEGGAARDGRIVQNDIIVEVDISLVGTQNFNA 575
 QY 660 ETLIRSMSTGKRMQILIVARRISKNEKSPSPGPELPTALDD--ERRRISHS 717
 DB 576 TVLR-----NKKGNVRFVGR-----EKPCQVSEVAKQLSQTLEORRRELLER 620
 QY 718 LYSIGIEGLDESPRNALSRINGESGKTLQSLTYNMPDODTIVITIEDRLPVLPHLSDS 777

DB 621 HYAQYDADD-----ETGETATDEE-----DEVEGPILP 649
 QY 778 SSSSHDVGVVTADAGTMAKAA--ISDSADCSLSPDVPVLAFOREGCRGRSMSEKRTQ 835
 DB 650 -----GGDMAIEVEFELPENEDMFSDDL-----TSKLISH 680
 QY 836 FSDASQDLFVTKRKSMDLGIADETKINTVDQKAGSPSDVGBSLGKSSLSLSIQ 895
 DB 681 FKEL-QIKHATVE-----AEIQKLT-----KLQA 704
 QY 896 AVAEVTLNDIPFHPRPRIINGRCNESFRAIDKSYDKPAVDDDEGMTLEEDTERS 955
 DB 705 AE-----NEKVAEIEKKNQLOQNIENKENVKL----- 733
 QY 956 SMSGRESVSTASDPSHSLERQMNQNGDKDTRKDKTKG----- 997
 DB 734 -----ESYWIEAQTLCHTVNEHLEKETOSYOALEKKYNAKKALINDFOQRELFIRQEV 788
 QY 998 EKKKDRDEKDKMKAKKGMKGLGDM---FR-----GKHKD 1032
 DB 789 EKKKLEVEKALIVEVQGLVRIKDLEAEVFRLLKQNGTQVNNNNNIFERRSPSPETVSKG 848
 QY 1033 DKIE--KTGKIKIQESFTS-----EERIRMKQOEKRIQAKTEFPEKQARERDYAEI 1083
 DB 849 DTMENVEVKQTSQDGLSDOLNVAVPETERLDSKALKTRAAQLSVNRQRPTFR----- 903
 QY 1084 QDFHRTGCDDELMYGVSYSBGSMAIAR-----POSFRGHMDLTYOVKKR 1134
 DB 904 -----LYSVSTDEOSLERKNFTFNDDESPSTSSAD--LSGLGAEPTPG 949
 QY 1135 NSKRPVDSKSTSPSNHRIQRLROEFOAKODEVEDERRTTYFEDQWPNARPATQGR 1194
 DB 950 LSOALISSDES---DWI-----DDELTD-----GOSPK 977
 QY 1195 HSYVVEVQKROREBRESQOARQYSLSPROSKNASSVSODSWEQ-----NYSPE 1248
 DB 978 HTQS-----QSRAYHMSVQVSHWLGLSLDYVEFSAQNISGQLQDQNTKL 1031
 QY 1249 GFSAKENRYSSTIGCSNGLGCHFNARVMEETQSLNBOBRREQO--KKOPPE 1305
 DB 1032 GMTSSODRALVKKLKEKMSLE---KARKAQEKKEKQKRLKREDOBOMRKSXSE 1086

RESULT 7
 DLG4_MOUSE STANDARD; PRT; 724 AA.
 AC 062108;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Presynaptic density protein 95 (psd-95) (Presynaptic protein SAP90)
 DE (Synapse-associated protein 90) (Disks, large homolog 4).
 GN DLG4 OR DLG4 OR PSD95.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=DKA/2; TISSUE=Brain;
 RA Kohmura N., Yagi T.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
 CC - FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS (BY SIMILARITY).
 CC ON THE PRESYNAPTIC SIDE (BY SIMILARITY).
 CC - SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).

DR EMBL: D50621; BAA09297.1; -
 DR HSSP: P31016; 1BE9.
 DR MGD: MGI:1277959; D1gh4.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR SH3 domain; Repeat.
 FT DOMAIN 65 151 PDZ 1.
 FT DOMAIN 160 246 PDZ 2.
 FT DOMAIN 313 393 PDZ 3.
 FT DOMAIN 428 498 SH3.
 FT DOMAIN 534 724 GUANYLATE KINASE.
 SO SEQUENCE 724 AA; 80472 MW; 7EFC99ELFF90BA CRC64;

Query Match 4.1%; Score 291; DB 1; Length 724;
 Best Local Similarity 21.6%; Pred. No. 5.1e-07;
 Matches 142; Conservative 89; Mismatches 205; Indels 220; Gaps 26;

OY 238 ODEGTEDNSVPEVGHADTGLEIHPNS-----LDDMKL 274
 DB 15 ODED-----TPLEHSPA--HLPNQANSPIVNTDLEARGYELQVNGTEGEM 63
 OY 275 VVPNDGPELGHVVPFSARGRT--LG---LLVRLKRGKAKHEMLFRENDCIYVI 327
 DB 64 EETLERGNSGL--GFSIAGTDNPHIGDDPSITTKIIPGAAADRLVNSILFV 120
 OY 338 NDGDLNRFRFEOAHMFROAMRTPIWPHVVPANKEQYEQLSEKNNYVSFSPDQ 387
 DB 121 NEVDREVYTHSAVAELKEA----- 140
 OY 368 YIDNSVNSAGIHTVQRAPLNHPPEQIDSHSLRPHSAHSPKPPSAPASAPQNVFTTV 447
 DB 141 -----GSIVRLYVRRRP--PAEKI----- 158
 OY 448 SSGYNTKKIKRLNTOLKKGTEGLFESITS--RDVTIGGSAPVYKNILPRGAIQDGR 505
 DB 159 -----IEIKLIKPGKGLGFSIAGVGNQHIIPDINSITYTKILEGAAHKDGL 206
 OY 506 KAGDLRIEYNGVDLVKSGQEEVYSLRSTKMEGVSLVFRQ-----EDAFHPRLENAEP 560
 DB 207 QIGDKILAVNSYGLDEYHEDAVALKNT--YDVVYLKVAKPSNATVSDSYAPDITTSY 264
 OY 561 SOMQIPKREK-----AEDEDIVLTPDGTREELTF-----EVP-----LNDSSASL 601
 DB 265 SO-HLDNELSHSSYLGTDVPTAMPTSPRYSPPVAKDLLEGDIPREPRRIYHRSSTGL 323
 OY 602 GVSVAQNSKENHADLGIIVKSIINGAASKDGRLLVNDQLAVNGESILGKTNDAMET 661
 DB 324 GNINIGGEGE-----GIFISIIAGGRADLSGLRKQIISVNGVDLRNASHQAA 378
 OY 662 LRRSSTBGNKGMTOIIVARIKSCNELKSPGSPGPPLPIETALDD--RERRISHSLVS 720
 DB 379 LKNA-----GQVTYIIAQ-----YKPEEYSREFAKHIDRLQMLNMSLS 418
 OY 721 GIEGLDESPRNAALSRIMGESGKYO---LSPYVMPODDIVIIEDDLPLVLPRLSLDQ 776

DB 419 GRASLNSNKRREYI-RALEFDYDKTDCGFLSGALSPHGDVLIHVID-----ASDE 468
 OY 777 S-----SSSHDVGFTVA-----DAGTWAKAAISDSADCSLSPD-VDPVLAFO 819
 DB 469 EWMQARVRSDBETDDIGFIPSKRRYERREWSRLKAKDWSSSGSGREDVSLSYE 524
 RESULT 8
 ID DGL1_HUMAN STANDARD; PRT; 904 AA.
 AC Q12959; Q12958; (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE presynaptic protein SAP97 (Synapse-associated protein 97) (Disce,
 DE large homolog 1).
 GN DGL1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95024052; PubMed-7937897;
 RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
 RT Cloning and characterization of hdlg; the human homologue of the
 RT Drosophila discs large tumor suppressor binds to protein 4.1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
 RX MEDLINE-96338231; PubMed-8757139;
 RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
 RA Marfatia S.M., Chishti A.H., Liddington R.C.;
 RT "Crystal structure of a PDZ domain.";
 RL Nature 382:649-652(1996).
 CC - FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNIT. ASSOCIATES WITH PROTEIN 4.1.
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC - SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC - SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).

DR EMBL: U13897; AAA50599.1; -
 DR EMBL: U13896; AAA50598.1; -
 DR PDB: 1PDR; 23-JUL-97.
 DR MIM: 601014; -
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 KW SH3 domain; Repeat; Alternative splicing; 3D-structure.
 FT DOMAIN 224 310 PDZ 1.
 FT DOMAIN 319 405 PDZ 2.

FT DOMAIN 466 546 PDZ 3.
 FT DOMAIN 581 651 SH3.
 FT DOMAIN 714 904 GUANYLATE KINASE.
 FT VARSPLIC 669 680 EIPDMGSKGK->OSFNDKRRKKNLFSRRKFPYKMKDOS
 FT SEQUENCE 904 AA: 100354 MW: 878796 DBB0920D4 CRC64;
 ROETSDADQ (IN ISOFORM 2).

Query Match 4.1%; Score 286.5; DB 1: Length 904;
 Best Local Similarity 19.5%; Pred. No. 1,1e-06;
 Matches 207; Conservative 125; Mismatches 343; Indels 385; Gaps 37;

QY 179 FLKONTAGSPKTCRRKKE-----NYSLPDTSNMNOFORDNARSLSASHPMV 229
 DB 55 FEYEVILDNPKCIDSKSEPTQIPVNTWEISLP-----SSTVTSPLPSLS---PSV 105
 QY 230 GAWLEKODEDEGTEED-----NSNVEVGH 235
 DB 106 EKY---RYODETPPOEHIISQITNEVIGPELVHSEKNLSEIENHGEVSHSHISPIKP 162
 QY 256 ADTGLEHLPNPSL-----DDWAKLYEVDNDGSP----- 283
 DB 163 TEAVLPSPPTVPIVLPVPAENVTVIPIPOANPPVLVNTDSLPTVYNTDADYEX 222
 QY 284 -----LGIHVPPSANGR-----TLGLVLRLEKGGKAKHBNLFRNDCIVRINDG 330
 DB 223 EETLERNGSGISAGTGNPHIGDSSIFTKITTGAAADRLRVNDICILQVNEY 282
 QY 331 DLNRREFQOHMFRQAKRTPIIWEHVPAANKROYEOLQSSEKNYSSRFSPSDOYID 390
 DB 283 DVROYVTSKAVEALKEA-----GSIVRLYVKKRPVSE--- 315
 QY 391 NRSVNSAGLHTVQAPRALNPPPOIDSHSRPLSHAHPSGKPPSAPASAAQNVFTTVSSG 450
 DB 316 ----- 315
 QY 451 YNFKIKGRNLQIKKGTGIGSITS--RDVTIGGSADIVYKNILPRGAALODRLKAG 508
 DB 316 -----KIMEIKILPGKLGSLAGVGNHIGPDNSIYVTKILEGAAHKDKLOIG 368
 QY 509 DRIEYVNDLVGKSOEEVYSLRST-----KMEGTVALFVREQEDAFHRELN----- 557
 DB 369 DKLLAVNNVCLBEVTHEEAVTAKMNSDEVYIAKAKPISMW---NDGYAPPDITNSSSQ 425
 QY 558 -----AESQWQIKETKAEDEDIVLPDGTRELFTEVPLDNGSAGL 601
 DB 426 PVDNHHVSPSPFLGOTPASPARISPVSKAVLDDEITREPR-----KVYLH-RGSTGL 476
 QY 602 GVSVAGNRSKENHADLIGIFVKSIIINGGAASKDGRRLRVNDOLIAVNGESLICKTQDMMET 661
 DB 477 GNIYVGEGEGE-----GIFISPLAGPADLSGELRKGDRLITVNSVDLRAASHEDQAAA 531
 QY 662 IRRSMSTGNKRGMIQLVARISKMEIKSPSGSPGPELPIETALDD-RERRISHLYS 720
 DB 532 LKNA-----GOAVTVVAD-----YRPEYSRFKAKIIDLREQMANSISS 571
 QY 721 GIEGLDESPSRNAALSRIMGESGKYQLSPYVNMQDDTVIIEDDRPLVPLPHILSDOSSSS 780
 DB 572 GSGSLRTSOKKSLVYRAL-----FDYDT--KDSGLP-----SOGIANKF 608
 QY 781 SHDVGFTVADAGTWAKAALSDSADCSLDVDPVLAFOREGREGROSMSKERTKOPSDAS 840
 DB 609 FGDILIHVINSDEEMWQAR-----QVTPDGE-----SDLV 638
 QY 841 QLDVFKTRKRSMDLGIADETKINTVDOKAGSPSRDVP-----SLGLK---KSSSL 890
 DB 639 GVIPSKRRVRKK-----ERARKTV--KFNSTKIDKGEIPDDMGSKGLKLVHTSNASDS 689
 QY 891 ESLOINAEVTLNGDIPFR-----PRPRIINGRCNESFRRAIDKSYDK----- 935
 DB 690 ESSYRGQEEVLYSTE--PVNQOEYVNTRPYIILGPMKDRINDLLISEFPDKRGSCVPHHTR 748
 QY 936 PAVDDDEG-----METLEEDTESSRSRESVSTASDQPSHSLEROMNGNOEKGD 987

DB 749 PKRDYEDVDGKDRHFVTVSRQWEMKDIQENK-----FIAGQYNNHLVQTS 792
 QY 988 TDKKKQTKTKKKKKKKDKRKKRKKAKK-GML-----KGLGDMFEGHRRDKLTK 1037
 DB 793 VOSVREVAKGKHCILIDVSGNAIKRQLQALYPLISIFIKPMSMENTMBMNRRLTEQARK 852
 QY 1038 TGC--IKIQESTTSSEERMRMKOEORQATREFRERQA 1075
 DB 853 TFERAKMLEQETTEHTFTAIQVGTLEDILYNOYKOIIEOS 892

RESULT 9
 DLG4_RAT
 ID DLG4_RAT STANDARD; PRT; 724 AA.
 AC P31016; P97631;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Presynaptic density protein 95 (PSD-95) (Presynaptic protein SAP90)
 DE (Synapse-associated protein 90) (Discs, large homolog 4).
 GN Dlg4 OR PSD95.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
 RX MEDLINE=93040233; PubMed=7680343;
 RA Kistner U., Menzel B.M., Veb R.W., Cases-Langhoff C., Garner A.M.,
 RA Appeltauer U., Voss B., Gundelfinger E.D., Garner C.C.;
 RT SAP90, a rat presynaptic protein related to the product of the
 RT Drosophila tumor suppressor gene dlg-A.";
 RT J. Biol. Chem. 268:4580-4583(1993).
 RN [3]
 RP SEQUENCE OF 566-625 FROM N.A.
 RC STRAIN=WISTAR KYOTO; TISSUE=Vascular smooth muscle;
 RA Adams L.D., Werry I., Schwartz S.M.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS) OF 302-402.
 RX MEDLINE=96270509; PubMed=8674113;
 RA Doyle D.A., Lee A., Lewis J., Kim E., Sheng M., Mackinnon R.;
 RT Crystal structures of a complexed and peptide-free membrane protein-
 RT binding domain: molecular basis of peptide recognition by PDZ.";
 RL Cell 85:1067-1076(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
 CC SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
 CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
 CC POSTSYNAPTIC).
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

Query Match 4.1% Score 286; DB 1; Length 870;
 Best Local Similarity 20.9%; Pred. No. 1, le-06;
 Matches 196; Conservative 113; Mismatches 303; Indels 324; Gaps 39;

```

QY 400 HTVQAPRLNHPPE---QIDSHSRLLPHSAHSGKPPSPAPAPQVNFSTTVSS--GYNT 453
DB 34 HEV-RGPELVAVSEKNLSQIENVHGYLQSHIS--ELKASPAPILVNTDITDIPYNG 89
QY 454 KRIKRL-NIDKKGTDELGSIT--SRDVTIGSAPILYKNILPRGAIADGRLKAGDR 510
DB 90 TEIEVEFEETLEKNGSLGSLGAGTGNPHIGDPPGIFITKIIPGAAADGRLRNDC 149
QY 511 LIEVNGVDLVKSOEEVYSLRSTKMEGTSLVLFROEDAFHRELNAEPSOMQIPKETK 570
DB 150 ILRVNEVDVSEVSHSKAVEAL--KENGSIARLVYRRR----- 184
QY 571 AEDEDIVLTPOGTRFEFLTFEPPLN-DSGSAGLVSVGNRKNEN-HADLGFVKSIING 628
DB 185 -----RPLETVEYIKLFKPKPKGLGFSIAGVGNQHHPGDNSTIYTKIIDGG 231
QY 629 AASKDGLRVNDOLIAVNGESILKTNODAMETLRSMSTEGNKGMIQILVARRISKN 688
DB 232 AAKQGRGLVGDRLMLVNNYSLEEVTHEEVAALLKNTSEVYLYKGNPTTIY----- 283
QY 689 ELKSGSPGPELPJETALDDERRISHSLSYGIDLESFRRNALSRINGESKYO-- 746
DB 284 -MTDFYGPDP-----ITHS-YS-----PPEHNLIS--GNNTLTLYK 316
QY 747 -----LSPTVNMPODDVYIIEDDRLPVLPPI-----LSDSSSSSHDDVGFVTADAG 793
DB 317 TSLPSPISPGYSPYIPKHMIVDDY--TRPEPYSTVKNLCKDKRPSPH----- 363
QY 794 TWAKAAISDSADCSLSPDVPVLAFOREGFGROSMSEKTKOFSASQIDEPYKTKKSSM 853
DB 364 -----YSPVRCDS-----FLLSAPYSHY 382
QY 854 DGIADETKLNIVYDQKASPSRDVPSLGLKSSLSLESLQTAVAEVLNIGDIPFHRPP 913
DB 383 HIGLPLDSEMTSHSOHSTATRO--PSMTLDRAVSLG-----EPK 421
QY 914 RIRI-----GRGNESFRRA-----IDKSYDKPAVDDDEGMEYLE 949
DB 422 VVLRKSTGIGFNIYGDEGEGIFVFTLAGPADLSELQGDQILSYNGIDLNGASHE 481
QY 950 EDTSESSRSGRSEVSTASDPS-----HSL-ERONGNGEKGDKDKKDKTKGEX 999
DB 482 QAAALUKGAQVYTIIAOYQPEDYARPEAKIHDLREOMNHSSGSGSLRTNOKKSLYV 541
QY 1000 KK--DBDKEDKKAKKMGMLKGLDMFRGKHRRDKRIETGKIKIOESFTSEERI--- 1054
DB 542 RAMPFDIDKSDGSLPSOGLSFKYGDILHV-INASDEEMQARVMEGD--SEMGVIPS 598
QY 1055 -RMKQDERTIQAT-----REFRERQARERYATLOPF 1086
DB 599 KRVERKERARLKTIVKNAKPGVYDSKSFNDRKKSFISRKFPYKKNKEOSEQETSP 658
QY 1087 HRTFGCDELMATGVSSESMALNARQSPRE-----GHMDALYAQVKPRNS 1136
DB 659 ER--GQEDLLI--SYE-----PVTRQELNTRPVIILGPKMDRI----- 693
QY 1137 KPSVDNSNRSTPSNHRDIQRLROF-----QOAKODEVEDERRRTYSFEQWPANRPA 1189
DB 694 -----NDDLISEPDKFGSCVPHTRPKRDYEDGR----- 724
QY 1190 TOSGRHSVYEVQORQORQER--ESSQOARQYSS 1223
DB 725 ---DYHFVISRQOMEKDIQEHKFIETAGQYDNLTG 757

```

RESULT 11
 DLG3_HUMAN

DLG3_HUMAN STANDARD; PRT; 817 AA.
 Q92796; Q9UL18;
 01-NOV-1997 (Rel. 35, Created)
 01-NOV-1997 (Rel. 35, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Presynaptic protein SAPI02 (Synapse-associated protein 102)
 DE (Neurodoctine-DLG) (NE-DLG) (Discs, large homolog 3).
 CN DLG3 OR KIAA1232.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=9733263; PubMed=9188857;
 RA Makino K., Kuwahara H., Masuko N., Nishiyama Y., Morisaki T.,
 Sasaki J., Nakao M., Kuwano A., Nakata M., Ushio Y., Sawa H.;
 RT "Cloning and characterization of NE-dlg: a novel human homolog of the
 RT Drosophila discs large (dlg) tumor suppressor protein interacts with
 RT the APC protein.";
 RL Oncogene 14:2425-2433(1997).
 RN [2]
 RP SEQUENCE OF 330-817 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
 Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:337-345(1999).
 CC - FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
 CC SUBUNIT NR2B (BY SIMILARITY).
 CC - SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC - SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC - SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U49089; AAB61453.1; .
 DR EMBL; AB033058; BAA86546.1; .
 DR HSSP; P31016; 1BE9.
 DR MIM; 300189; .
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Pfam; PF00595; PDZ; 3.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00072; Gukc; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS00852; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR PROSITE; PS50002; SH3; 1.
 DR SH3 domain; Repeat.
 KW SH3 domain; Repeat.
 FT DOMAIN 130 217 PDZ 1.
 FT DOMAIN 226 311 PDZ 2.
 FT DOMAIN 379 465 PDZ 3.
 FT DOMAIN 503 568 SH3.
 FT DOMAIN 628 803 GUANYLATE_KINASE.
 FT DOMAIN 803 883 GUANYLATE_KINASE.
 FT CONFLICT 330 381 PRIMAEDPT -> AARREGAMERAKRFGSGIAMIIGS
 ASASAWRASQRMAMPFLASLRPGDA (IN REF. 2).


```

FT CONFLICT 450 454 MISSING (IN REF. 2).
FT CONFLICT 464 464 CD -> RK (IN REF. 2).
FT CONFLICT 474 474 D -> H (IN REF. 2).
FT CONFLICT 476 476 R -> P (IN REF. 2).
FT CONFLICT 478 478 A -> D (IN REF. 2).
FT CONFLICT 484 486 AAA -> LP (IN REF. 2).
FT CONFLICT 506 506 A -> S (IN REF. 2).
FT CONFLICT 569 569 H -> N (IN REF. 2).
FT CONFLICT 586 586 L -> Q (IN REF. 2).
FT CONFLICT 626 641 MISSING (IN REF. 2).
FT CONFLICT 639 639 K -> A (IN REF. 3).
FT CONFLICT 726 726 F -> L (IN REF. 1).
FT CONFLICT 733 733 E -> V (IN REF. 2).
FT CONFLICT 749 749 E -> V (IN REF. 1).
FT CONFLICT 756 756 L -> H (IN REF. 2).
FT CONFLICT 791 792 KR -> NG (IN REF. 2).
FT CONFLICT 794 794 T -> M (IN REF. 2).
SQ SEQUENCE 852 AA: 94934 MW: F8DA14AB9CF5B09 CRC64;

```

Query Match Beat Local Similarity 4.0%; Score 279; DB 1; Length 852; Matches 193; Conservative 112; Mismatches 311; Indels 298; Gaps 40;

```

OY 400 HTVORAPLNPPE-----QDSHSLRPHSAHPSGKPPAPAPQNVSTVSS--GYNT 453
DB 34 HEV-RGPVLVHSEKNSLQLENHGVLOSHTS---PLKASAPLIYVTDITDITPYVNG 89
OY 454 KIKGRRL-NIQLKKGTEGEGFSIT--SRDVTGGSAPLIYVNIIPRAAIDGRLKAGDR 510
DB 90 TELIEFEETTLERNGSGIGFSIAGTDNPHIGDPPGFIETIKIIPGAAEDRLVAVNOC 149
OY 511 LIEVNVDLVYKSGOEVSLSLSTMEGVSLVFEQEDAFHPELNAEFSQMDIKREK 570
DB 150 ILKNEVSEVSHSKAVALE--KEAGSIYRLVYRR----- 184
OY 571 AEEDIVLPPDGTREFLEFEVPLN-DSGSAGLGVYVGNRSKEN-HADGIFVKSITNG 628
DB 185 -----RLEETVEILFRGPKGLFSIAGVGOMHPIPGDMSIVTKIIDGG 231
OY 629 ASKSGRLRVNDOLIAVNGESLIGKTQMDMETLRSMSTEGNKGMIQIIVARRISCN 688
DB 232 AAQKGRQVQDRLMANNYSLEEVYHEKVALKNKTSO-----VYLKVGKPT 280
OY 669 E-LKSPGSPGPELPIETALDRERRISLSYSGIEGLDESPHNALSLIMESKTYQ 746
DB 281 TIYTDVYPPD-----ITHS-YS-----PPMNHLS--GNNGTLE 314
OY 747 -----LSPTVMPODDVTIIEDRLVPLPH-----LSQSSSSSHDDVGVVAD 791
DB 315 YKSLPSIPSGRSPITPKHMLVEDDY--TRPEPVYSTVKNLCKDKPASPRLH----- 363
OY 792 AGTWANAALSDADSLSPDVVLAFOREFGNOSMEKRTQFSDASQDLDFVTKRCK 851
DB 364 -----YSFVECDKS----- 851
OY 852 SMDLGIADETKINYYDOKAGSPSRDVPGLGKSSLESLOTANAETLVNGDIPFRHP 911
DB 381 IYHGLDLPDSMTSHSHSTATRO--PSVTLQRAISLEG-----FLSLSTPP 380
OY 912 RPRIR-----GRGCHESPRAA-----LDKSYDPAVDDDEGMEP 947
DB 420 KKVVLHGSGTGLGFNIYGEDGIVSTLLAGGPADLSGELORDQILTSVNGIDLNGAS 479
OY 948 LEEDTESSSGRESVSTASDPS-----HSL-FRONGNGDEGDKTRKKRTGK 997
DB 480 HEQAALAKAGQVTYIIAQYQPEYARPEAKIHLRQNMNHSMSGSSSLFTNKRSL 539
OY 998 EKKK--DROKEDKMAKAKGKMGKGLDMFRGKRKDKLEKTKIKIDESTSEERI- 1054
DB 540 YVAMDFDYKSDSLPQSGLSFKGGLIYH-IVASDEWMAQARVILDGD--SEMGVI 596
OY 1055 ---RMQOEORIQAKTRFRRQARRENDYAEIODFHFTFCDELMYK-GVSYSG--SM 1108

```

```

DB 597 PSKRVERKERARLTKVFNAKPGVIDSKGDIP-----GLQDD---CYGKTLRGEDL 647
OY 1109 ALNARPOSPE-----GHMDALYAQYKPKPRNSKSPVDSNNTSNDRIORL 1158
DB 648 ILSEYPTROELNVTTPVILLSPMDRI-----NDLISPEP 684
OY 1159 QEF-----QAKODEVEDRRRTYSFEQPPNAPRATOSGRHSVVEVQMDROEER 1211
DB 685 DKFGCVPHITRPRKRYEDGR-----DYHFVISEOMKEDIQEK 725
OY 1212 --ESSQOAOORYSS 1223
DB 726 FIEAGGYNDNLVGT 739

```

RESULT 13

```

ID DGL1_RAT STANDARD; PRT; 911 AA.
AC 062696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Presynaptic protein Sap97 (Synapse-associated protein 97) (Discs, large homolog 1).
GN DGL1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid:10116;
RN (1);
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95198112; PubMed=7891172;
RA Mueller B.M., Kistner U., Ven R.W., Cases-Langhoff C., Becker B.,
RA Gumbelinger E.D., Garner C.C.;
RT Molecular characterization and spatial distribution of Sap97, a
RT novel presynaptic protein homologous to Sap90 and the Drosophila
RT discs-large tumor suppressor protein.
RT J. Neurosci. 15:2354-2366(1995).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN
CC CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: IN CULTURED 784 CELLS, IT IS RESTRICTED TO
CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT
CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
CC -1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
CC SYNAPSES IN OTHER BRAIN REGIONS. SAP97 IS FOUND IN AND ALONG
CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
CC VARIETY OF EPITHELIAL CELLS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL; U14950; AA79976.1; .
DR HSSP; 012959; 1PBR.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin.1.
DR Pfam; PF00595; PDZ.3.
DR Pfam; PF00018; SH3.1.
DR SMART; SM00072; GUKC; 1.

```

```
DR SMART: SM00228; PDZ: 3.
DR SMART: SM00326; SH3: 1.
DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ: 3.
DR PROSITE: PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 224 310 PDZ 1.
FT DOMAIN 318 404 PDZ 2.
FT DOMAIN 465 545 PDZ 3.
FT DOMAIN 580 650 SH3.
FT DOMAIN 721 911 GUANYLATE KINASE.
FT DOMAIN 527 530 POLY-ALA.
SQ SEQUENCE 911 AA; 100570 MW; 18CEBD31DDOCAPRB CRC64;

Query Match 4.0%; Score 278.5; DB 1; Length 911;
Best Local Similarity 20.7%; Pred. No. 2.7e-06;
Matches 190; Conservative 129; Mismatches 344; Indels 235; Gaps 35;

OY 359 PAANKROYEOLSQSEKNNYSSRFSPDSQYIDNRSVNSAGLHTVORAPRLNHPPEQIDSH 418
D 103 PPEYKRYQ-----DEVLPSEKISPO--VPNEVLGPVLVHVSSEKLSLEINVGFSH 154
OY 419 SRL-----PHSAHPGCKP-----PSAPASAPQNVFTVSSGYNTKTI 456
D 155 SHISPIKPLEAVPPSSPIVPTPALPVPAPESPVPVLPSTPOANPPVLVWTSLETPTYVN 214
OY 457 GKRLN-----IOLKKGTEGLGFSIT--SRDVIIGGAPVIYVNIIPRGAIODGLKAGD 509
D 215 GTDADYEIEETLEKNGSLGFSIAGTDNPHIGDSSIFITKITTGGAAADGRLRND 274
OY 510 RLIEVNGVDLVGKSGOEYVSLSTRKMEGTVSLVRFQEDAFHRELNAEPQOMQIPKET 569
D 275 CILRVNADVDRVTHSKAVEAL--KEAGSIVRLVYKRRKAR----- 314
OY 570 KAEDDIYLVFDGTRFEPLFEVPLNDSSAGIGVSKGNRSEN--HADLIGFVKSTINGG 628
D 315 --KNEIETLI-----KPKPGIGFSIAGVGNGHLPGNSIVYTRITIEGG 356
OY 629' AASKDRLRVNDQLIANGESILGKTNDAMETLRRSMTEGNGMIOIIVARRISK-C 687
D 357 AAHKGKIQIDGKILAVSVCLSEVTHEAVYALK-----NTSDFYLKAAKPTSMYI 409
OY 688 NELKSPGSPPEPELPIETALDRERRISHISLYSGIEGDESFRNAALSIR-IMGES--- 742
D 410 NDGYAP-----PDI-TNSSSQSVNDHVSPPSYLG--QTPASPARVSPISKAVLGDDETR 461
OY 743 -----GKQLSPYVNMQDDVITIEDRLPVLPHLSQ-----SSS 779
D 462 EPRKVVLRHGSTGLGFNTVGEGEGIFISFLIAGGPADLSGELRKGRRIISVNSVDLRA 521
OY 780 SSHDVGVEVTADAG-----TWAKAISDSADCSLSPDVPVLAFOREGEGFR 825
D 522 ASHEDAAAALNAGQAVITVAQYRPEEYSRFAKIHDLRETMANSSVS-----SGSGS 574
OY 826 QSMSEKRTKQFSDASQDLFVTKTRKSKMDLGIADETKLTNYDDQKAGSPSRDY----GPS 881
D 575 LRTSQKRLYV--RALFDYDKTDS-----GLPSQGLNFKFGDI 611
OY 882 LGAKSSSLESIQTAFAVAVTLNGD-----IPFHRPRRIIRGRCNESPRAIDKSDK 935
D 612 LHVINASDDEWMO--ARQYTPDGESDEGVVIPSRRVKEKRAARKTYFNS--KTRGDK 667
OY 936 PAVDD--DEGMETLEEDTEESSRGRSVSTASDOPSHLSEROMNGNOE-----K 984
D 668 GEIPDKMSKGLKHYTASNABSESSYHEYGCGSKGOEEVVLSEYVNOQEVVYTPVYL 727
OY 985 GDKTRKDKTKGE-----KKRDRDEKDK-----MKAKGMLKGLG-----D 1022
D 728 GPMKDRVNDLLESPDFKFGSCVPTTRPKRDYEVDRYHVTYRREQEKDIQEHKFE 787
OY 1023 MFRFGKHKRDKIE-----KTKGIKIQESTSEERTIMQO-----EDERTIA 1065
```

```
D 788 AGQYNHLYGTSVQSVRAVAERKGCITLDVSGNAIKRIQIADLYPISIFIKRSMENIME 847
OY 1066 KTRFERERAR---ERDYAEIODFHRFCGDDLEMTGVSYSSEGMALNARQSPREGHM 1122
D 848 MKKRLTEDQARKTFERAVALEQEFTEHF-----TAIVQGIT 883
OY 1123 MDALYAQVYKRPNSKRPSP 1140
D 884 LEDIYNQVYKQIEEQSGP 901

RESULT 14
ATTRX_HUMAN
ID ATTRX_HUMAN STANDARD; PRT; 2492 AA.
AC P46100; P51068; O15886; O9NTRS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator Attrx (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATTRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND VARIANT S-1860,
RP AND VARIANTS ATTR-X.
RX MEDLINE=97123494; PubMed=8968741;
RX Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RX Glibons R.J.;
RT "Attrx encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the Attr-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RX Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RX Coliaux L., Schwartz C., Fontes M.;
RT "determination of the genomic structure of the XNP/ATTRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RX Statton C.L., Dabovic B., Gulisano M., Gecz J., Broccoli V.,
RX Giovannuzzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RX Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050; Villard L., Statton C.L.,
RX Gecz J., Pollard H., Consalez G., Villard L., Higgs D.R.;
RX Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKCI in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATTR-X.
RX MEDLINE=95211835; PubMed=7697714;
RX Glibons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (Attr-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RX Pearce A., Chapman J.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBD databases.
RN [7]
RP E2R2 BINDING.
```

RX MEDLINE=96167853; PubMed=9499421;
 RA Cardoso C., Timst S., Villard L., Khrestchatsky M., Fontes M.,
 RA Colleaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RT domain of the human EZH2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 RN [8]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE=20040663; PubMed=10570185;
 RA McQuowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 RN [9]
 RP DISEASE.
 RX MEDLINE=20213147; PubMed=10751095;
 RA Villard L., Fontes M., Ades L.C., Gecz J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 RN [10]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE=97196774; PubMed=9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RT without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 RN [11]
 RP VARIANT JM GIN-2131.
 RX MEDLINE=96224392; PubMed=8630485;
 RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,
 RA Munlich A., Lyonnet S.;
 RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 RN [12]
 RP VARIANTS ATR-X.
 RX MEDLINE=97467722; PubMed=9326931;
 RA Gibbons R.J., Bachoo S., Picketts D.J., Affinos S., Asenbauer B.,
 RA Berghoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Stanley S.F.,
 RA Higgs D.R.;
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 RN [13]
 RP VARIANT ATR-X LEU-246.
 RX MEDLINE=20123062; PubMed=10660327;
 RA Fichtera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 RN [14]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE=99347960; PubMed=10417298;
 RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
 RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 RN [15]
 RP VARIANT CMS THR-2050.
 RX MEDLINE=99326061; PubMed=10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.;
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).

RN [16]
 RP VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE=99219535; PubMed=10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belongue J.,
 RA Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichtera M.,
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).
 RN [17]
 RP VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
 RX MEDLINE=20451413; PubMed=10995512;
 RA Wada T., Kubota T., Fukushima Y., Saitoh S.;
 RT "Molecular genetic study of Japanese patients with X-linked
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEKIN V IN A CALCIUM AND
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 CC INTERACTING WITH HP1.
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1, 2, 3, 4 (SHOWN HERE) AND 5;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBICITOUS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
 CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
 CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
 CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
 CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
 CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
 CC SYNDROME (CMS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
 CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH
 CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME
 CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
 CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
 CC MICROGENITALISM AND EARLY DEATH.
 CC -1- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS
 CC SYNDROME (SFM). CLINICAL FEATURES INCLUDE SEVERE MENTAL
 CC RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND
 CC BILATERAL CRYPTORCHIDISM. DUE TO THE CLINICAL OVERLAP WITH ATR-X
 CC SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SFM, MIGHT
 CC BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H
 CC INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED
 CC MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE
 CC MENTAL RETARDATION WITH SPASTIC PARAPLEGIA, MICROCEPHALY, SHORT
 CC STATURE AND CRYPTORCHIDISM.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U72937; AAB49970.2; -;
 CC DR EMBL: U72938; AAB49971.2; -;
 CC DR EMBL: U72935; AAB40698.1; -;
 CC DR EMBL: U72904; AAB40698.1; JOINED.

Query Match 3.5%; Score 246; DB 1; Length 2492;
 Best Local Similarity 17.7%; Pred. No. 0.00036;
 Matches 241; Conservative 217; Mismatches 516; Indels 384; Gaps 53;

```
OY 114 SAFOPYQATSELETP--SVLRAMP LHVRRSSD-ALIGLSTSVDSNFS-----SEEP 165
DB 415 SEFRAMDVNNKEKTKREKHVIDAKETRAKGEKPCALEKNDISSEAKLSTKROYDSEIM 474
OY 166 SKRNTFMASTTAGFLKONTAGSPKTCDDRKDDNYRSLPRDTSNWNQOROMARSLAS 225
DB 415 HONVTEQORT-----NKSTGGEHKSDDRKEEPQY-----EPANISEDLDMDIV--SVPS 523
OY 226 HPWVKMLEKQEDJEDNDSRVEPYGHADTGLEHIP-----NFSLD----- 270
DB 524 VP-----EDJFENLETA MEVQSVYDHOQDSSGTEQEVESSVKLTNISKNNRGKIKX 577
OY 271 -----MKLVYVPDNGPL---GHHVYFPSARGRTGLLYKRLKEGG---KAEH 314
DB 578 TTAQYTKELLYVLTTPVSPJNSPIKADCEVPODKGYKSCGLNPK-LEKCGIGQENSN 636
OY 315 ENLFEENDCIYVINDGLRNRRFEQAOHMFQAMRTPIIMFVYPAANQEYQOLQSQSK 374
DB 637 EHLV-ENEVSLEESDLR-----RSPRYKTPPLRPPIETNPVTNSNDECEETVAKOK 690
OY 375 -----NNYSSRSP-----DSQYIDNRSVNSAGLHTYOAPRLNHPPEQ 414
DB 691 LSVPRKKDRSSDSDALDNPKNLPPKSKQSEYVDQNSDSEMLALLKGVSRMSSSS 750
OY 415 IDSRLSHSAHPGKPPSPASAPQWFTTSSGYTKTKGKRLNLOLQKGTBGLGFS 474
DB 751 -DTDINEIHTN-----KTIYDLTKQAGKDK--GKR---KRKSTSGSDP- 790
OY 475 ITRSDVTIGGSAPIYV-----KNILPRGAIIODGRLLKAGDRLIEVNGVDLVGSGQEEVVL 530
DB 791 ----DTKGKSKSSITISKKQOTQSESSNDSELEK-----EIKSMKGTGA----- 834
OY 531 LKSTMEGTSLVLERODAFHR-----ELNAPESOMOIPEETRAEDD 575
DB 835 -RTTKKRIPNTKDFSSDEKSKKGMNQHKNLKTSQEGSSDAERQOEBETSSAAG 893
OY 576 IYLTEDGREFITEFVPLNDSSAGL-GYSVKGANSKEHNLGIFVSIINGAASKG 634
DB 894 TV-DKDTIMELDRLLPKQOASASTDGD--KLSGKQSTSTSLERKVAETKESKEL 949
OY 635 RL-----RVNDOLIAVNGESLGTNODAMETLRRSMSTGCKNKGMLQILVARISKNEL 990
DB 950 KTKTKKKVODGLSDI-AEKFLKKDSDSEDDKQSKGTEKKKPPSPFKKVVIMEQO 1008
OY 691 KSPGSPPELP-----IETALDDREERI-----SH 716
DB 1009 YESSSDGTREKLEPEREICHFPKGIQIKNGTGDGKSKKINDKTSKKKDELSDVAEKET 1068
OY 717 SLYSIGIEGIDESPNNALSR-----IMGESG-----KYOLSPTVNNPQDDT 758
DB 1069 GKGDCSDSEDKSKNGAYGREKKRCKLLGKSSRRKODCSSSDTEKYSM-----KEDG 1121
OY 759 VIEEDRLPYLPHLSDDSSSSSHDDVGFVTADAGTMAAALISDSDCS-----LS 809
DB 1122 CNSSDKRLRI--ELRERNLSSKRNTEIGSSSS-----SDAESSEDNKKKKKQRTS 1173
OY 810 PDVVDVLAFAOREGFGROSKSEKRTQFSDASOLDVEKTRKSKSMLGLIDETKLTAVDDO 869
DB 1174 SKKKAIVYKERNLRTSTKTKKQADITSSSSD--IEDDDQNSIGEGSDEKIKIPVTEN 1232
OY 870 KAGSPSRDVGPSGLKSSSLESLOTAAVEVTLNGDIPHRPRPRIIRGRCGNESFRRAI 929
DB 1233 LV-----LSHTGFCOSSGDEALSKSV-PYVDDDDDDNDNDENRIAK-KMLEELKAML 1284
OY 930 DKSVDKPAVDDDEGMETLEEDTESSNRSGRSVSTASDPHSLERONGNOEGKDDT 989
DB 1285 SSDEGSSSDPEEGKRTGKONEEP-----GEEAKNQVN 1321
OY 990 RKKDTGKEKKKDRDKEKDKMAKKGMLGLDMFRFGKHRRDDKIEKTGKIKIOESFVS 1049
DB 1322 SSDDSDSESKKPRYRH-----RLNLR-----KLTVSGGEGG 1353
OY 1050 EEBRIMKOEOERIOAKTRERERQARENDYAEIODFHRTFGCDDLEMTGVSSTEGSMA 1109
```

```
DB 1354 EEEK-----TKPEHEVKGRRNRKVSSEDE-----DSDFQSSGVSEEVSESE 1397
OY 1110 LNAPOSPEGHMDALYQVKKPRNSKPSVDNSSTPSNDRIOQLROEQOAKODD 1169
DB 1398 DQORPRT-----RSKAKAELENO-----RSKQKKRRRIIVQED 1433
OY 1170 VEDRRRTYSFEQPPWNPARNATQSGRHSVSEVQOMORQOEERSSQOARQYSSLPROSR 1229
DB 1434 SSEKNSNSEE-----DEEKEEEEEEEEEEDBNDSKSPGKR 1477
OY 1230 KNAASVSDSMEOANTSPEGCFOSAKENPRYSSTQGRNLTGCHGNARYMETQLELQ 1289
DB 1478 KIRIKLKND-----KLIRETONALKE 1499
OY 1290 EORRQOKMKOPSEGPSNYSYKKVODPSYAPKGP 1327
DB 1500 EEBRRRIARERERE--KLREVIETEDAS-PIKCP 1532

RESULT 15
NEBL_HUMAN
ID NEBL_HUMAN STANDARD; PRT; 742 AA.
AC Q9DJL8; 076059;
DT 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DR 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurabin-1 (Neural tissue-specific F-actin binding protein 1)
GN (Protein phosphatase 1 regulatory subunit 9a) (Fragment).
OS PPP1R9A OR KIAA1222.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirotsawa M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
RN [2]
RP SEQUENCE OF 512-742 FROM N.A. (SHORT ISOFORM).
RA Hinds K., Tin-Wollam A., Becker M., Stromatt C.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: BINDS TO ACTIN FILAMENTS (F-ACTIN) AND SHOWS CROSS-
CC LINKING ACTIVITY. BINDS ALONG THE SIDES OF THE F-ACTIN. MAY BE
CC INVOLVED IN NEURITE FORMATION. INHIBITS PROTEIN PHOSPHATASE 1-
CC ALPHA ACTIVITY (BY SIMILARITY).
CC -1- SUBUNIT: POSSIBLY EXISTS AS AN HOMODIMER, HOMOTRIMER OR AN
CC HOMOTETRAMER. INTERACTS WITH F-ACTIN. PROTEIN PHOSPHATASE 1 (PP1),
CC NEURABIN-II AND P70-S6K (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ALSO FOUND IN THE SOLUBLE
CC SYNAPTOSOMAL FRACTIONS (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM. MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: INTERACTS WITH P70-S6K VIA ITS PDZ DOMAIN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-bb.ch).
CC
DR EMBL: AB033048; BAA6536.1; -
DR EMBL: AC004022; AAC35294.1; -
```

DR HSP: 012923: 3PDZ.
 DR MM: 602468: -
 DR InterPro: IPR001478: PDZ.
 DR InterPro: IPR001660: SAM.
 DR Pfam: PF00595: PDZ; 1.
 DR Pfam: PF00536: SAM; 1.
 DR SMART: SM00238: PDZ; 1.
 DR SMART: SM00454: SAM; 1.
 DR PROSITE: PS50106: PDZ; 1.
 DR PROSITE: PS50105: SAM_DOMAIN; 1.
 KW Neurogenesis; Actin-binding; Cytoskeleton; Coiled coil;
 KW Phosphorylation; Alternative splicing.
 FT NON_TER 1
 FT DOMAIN 148 236 PDZ.
 FT DOMAIN 632 695 SAM.
 FT DOMAIN 241 271 COILED COIL (POTENTIAL).
 FT DOMAIN 314 468 COILED COIL (POTENTIAL).
 FT DOMAIN 677 734 COILED COIL (POTENTIAL).
 FT DOMAIN 79 122 INTERACTS WITH PROTEIN PHOSPHATASE 1-
 FT DOMAIN 96 103 GAMMA (BY SIMILARITY).
 FT DOMAIN 91 94 INTERACTS WITH PROTEIN PHOSPHATASE 1-
 FT DOMAIN 473 477 ALPHA (BY SIMILARITY).
 FT MOD_RRS 104 104 INTERACTS WITH TGN38 (BY SIMILARITY).
 FT MOD_RRS 104 104 POLY-GLU.
 FT VASPLIC 604 611 POLY-ASN.
 FT VASPLIC 611 611 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
 SQ SEQUENCE 742 AA; 84341 MW; A464IDCA457FADA0 CRC64;

Query Match 3.4% Score 238.5; DB: 1; Length 742;
 Best Local Similarity 18.8%; Pred. No. 0.00017;
 Matches 162; Conservative 106; Mismatches 254; Indels 339; Gaps 32;

QY 518 DLVKRSQEEVSLRSTMEGTSLVFR--QEDAFHPRRLNAPESQOMQPKETKAEDD 575
 DB 85 EIVGLPEEBEIPANRKIKF--SSAPIKVENTSNEDYDRNDEVPAVASAFYELEKREK 143
 QY 576 IVLTDPDGTRELFEEVPLNDSSA----GLGVSYKGNRSKENHADLGIFVKSITNGAAS 631
 DB 144 LELPP-----VELKEDDEGLISITIGMGVADAGLEK-----LGIFVKTVEGGAQ 190
 QY 632 KDGRLRYNDOLIAVNGSLGKTNDAMETLRRSMSTEGNKRGMOLIVARRISKNELK 691
 DB 191 RDGRIQVNDQIVEYDGLSVGTQNFATVLR-----NTKGNRVYIGR-----E 235
 QY 692 SPGSPPEELPIETALDD--RERRISHLSYSGIEGLDESPRNALSRIMGSGKYOLSP 749
 DB 236 KPGGVSEVAQLISQTLRQERQRELQOHVAQYDADD-----ETGEVATDE 282
 QY 750 TVNRPQDDTYIIEDRLPYLP-----PHLSQSSSSSHDDYGVYADAGTWAKA 798
 DB 283 E-----EDEVGPVLPESDMAIEVPELPENEDMFSPSELD----- 316
 QY 799 AISDADCSLSPDVPVLAIFORREGFGQSMSEKRTKQFSDASOLDPVKTRKSRMDIGIA 858
 DB 317 -----TSKLSHKFEL-QIKHAYTE-----A 336
 QY 859 DETKLNTVDDOKAGSPRDVGPISGLKSSLSLESIQTAVAETVLNGDIPFHRPRRIIRG 918
 DB 337 EIQKLKT-----KLQAAE----- 349
 QY 919 RGCNESRAAIDKSYDPAYVDDDDGEMETLEEDTEESSRGRESVSTASDOPSHLEROM 978
 DB 350 ---NEKYRWELEKTLQOONIEENKERMKL-----ESYWIETQILCHTVNEHL 394
 QY 979 NGNOEGDKTDRKKDKTKG-----EKRRDRXKEDKKAKKMKMLKGL 1020
 DB 395 KETOSQYQALEKTKNKKALKIKDQOKELDFIKQOEAKRKIEDLEKAHLVEVQGLQVRI 454
 QY 1021 GDM---FR-----FGHRKDKDKIEK--TGK--IKIOESFTS---- 1049
 DB 455 RDLEAEVFRLLKQNGTQVNNNNNIFERRTSIGEVSKGDTWMENLDGKQTSQDDGLSQDLNE 514

QY 1050 ---EERIIRMKOEQRIOAKTREFRERQARERDYAEIQDFHRTPGCDELMYGGVSSVEG 1106
 DB 515 AVPETERLDSKALKTBRAGLSYKNRRQRPSTFR-----LYDSVSTDG 556
 QY 1107 SMALNAR-----POSPREGHMDALYAOYKPKPRNSKPSVDSNRSTPSNHDRI--- 1154
 DB 557 EDSLERRKFTFENDFSPSTSSAD-LSGLGAEPKTPGLISGLALSDESLDMIDDEILDD 615
 QY 1155 -----QRLRQEF--QO----- 1163
 DB 616 QSPRKHSCCNRAVQEMSVOQVSHWMLMSLINEQYVSFSAONTTGEOLLQDGNKLKALG 675
 QY 1164 --ANQDEVEDRRRTYSFEQFPNARPATOSGRHSVSEVQMRORQREESSQQAQROY 1221
 DB 676 MTASQDRAVY-KKKLKEKKMSLEKARKAQE-----KMERQREKLRKKEQEQMQRK 724
 QY 1222 SSLRQSRKNASSVSQDSWEQ 1242
 DB 725 S---KTEKMTSTYAEAGAGEQ 742

Search completed: July 24, 2002, 14:34:04
 Job time: 250 sec